

**This Page Is Inserted by IFW Operations
and is not a part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- **BLACK BORDERS**
- **TEXT CUT OFF AT TOP, BOTTOM OR SIDES**
- **FADED TEXT**
- **ILLEGIBLE TEXT**
- **SKEWED/SLANTED IMAGES**
- **COLORED PHOTOS**
- **BLACK OR VERY BLACK AND WHITE DARK PHOTOS**
- **GRAY SCALE DOCUMENTS**

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2004, 16:23:59 ; Search time 48 Seconds
(without alignments)
3121.623 Million cell updates/sec

Title: US-09-964-238-2
Perfect score: 4991
Sequence: 1 MLTQAGNRKFKTECGKAF.....DGKMKTSKDHEDNEDMG 944

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

1:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4991	100.0	944	20	AAW83318
2	2612	52.3	533	22	ABR21853
3	2612	52.3	533	22	AAW57277
4	2612	52.3	533	22	AAW69682
5	2612	52.3	533	22	AAW30015
6	2612	52.3	533	23	ABG39298
7	1657.5	33.2	1125	24	ABU57636
8	1005	20.1	190	22	ABG48956
9	1005	20.1	190	22	ABG51171

10	1005	20.1	190	22	ABR28955	Peptide #1606 enco
11	1005	20.1	190	22	ABR31134	Peptide #3785 enco
12	1005	20.1	190	22	ABR34123	Peptide #1629 enco
13	1005	20.1	190	22	ABR36333	Peptide #3839 enco
14	1005	20.1	190	22	ABR19564	Protein #1563 enco
15	1005	20.1	190	22	ABR21680	Protein #3689 enco
16	1005	20.1	190	22	AAW54913	Human brain expres
17	1005	20.1	190	22	AAW57099	Human brain expres
18	1005	20.1	190	22	AAW67294	Human bone marrow
19	1005	20.1	190	22	AAW69491	Human bone marrow
20	1005	20.1	190	22	AAW15131	Peptide #1565 enco
21	1005	20.1	190	22	AAW17330	Peptide #3764 enco
22	1005	20.1	190	22	AAW27586	Peptide #1623 enco
23	1005	20.1	190	22	AAW29830	Peptide #3867 enco
24	1005	20.1	190	22	AAW02872	Peptide #1554 enco
25	1005	20.1	190	22	AAW05013	Peptide #3695 enco
26	1005	20.1	190	23	ABG36941	Human peptide enco
27	1005	20.1	190	23	ABG39120	Human peptide enco
28	817.5	16.4	747	22	ABR67181	Human peptide enco
29	817.5	16.4	1054	22	ABR57993	Drosophila melanog
30	717	14.4	141	19	AAW58368	Human CN238.1 prot
31	713	14.3	141	19	AAW58369	Human secreted pro
32	380.5	7.6	1252	22	AAW79739	Human protein SEO
33	379	7.6	1472	22	AAU31578	Novel human secret
34	374	7.5	1230	22	AAU30831	Novel human secret
35	371.5	7.4	1212	22	ABG00339	Novel human diagno
36	357	7.2	1342	22	ABG01726	Novel human diagno
37	356.5	7.1	1042	23	ABG36316	Human polypeptide
38	356.5	7.1	1395	23	ABG95054	Human ovarian canc
39	355.5	7.1	1051	23	ABG36314	Human ovarian canc
40	354	7.1	878	22	AAU29528	Novel human secret
41	350	7.0	1396	22	ABG08226	Novel human diagno
42	349.5	7.0	923	23	ABP69037	Human polypeptide
43	349.5	7.0	1520	22	ABG27130	Novel human diagno
44	346.5	6.9	973	22	ABG21022	Novel human diagno
45	341.5	6.8	1939	23	ABP62852	Human polypeptide

ALIGNMENTS

RESULT 1					
AAW83318					
ID	AAW83318	standard; Protein, 944 AA.			
XX	AAW83318;				
XX	20-MAR-2003 (updated)				
DT	01-MAR-1999 (first entry)				
XX	Mouse SMAD interacting protein S1P1.				
DE	S1P1; SMAD interacting protein 1; mouse; embryogenesis; diagnosis.				
KW	differentiation; cancer; neural disease; immune disease; therapy;				
KM					
OS	Mus sp.				
XX					
FH	Key	Location/Qualifiers			
FT	Region	166..216			
FT		/note="Smad binding region"			
XX					
PN	W0985512-A2.				
XX					
PD	10-DEC-1998.				
XX					
PF	28-MAY-1998;	98WO-EP03193.			
XX					
PR	02-JUN-1997;	97EP-0201645.			
XX					
PA	(VLA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOL.				
XX	HuyLebroeck D, Remacle J, Verschueren K;				
PI					

XX WPI, 1999-045732/04.
DR N-PSDB; AAV72502.

XX New SMAD interacting proteins obtained by a two hybrid screening
PT assay - useful for diagnosing developmental disorders and e.g.
PT cancer, immune and neural disease

XX Claim 7; Page 32-34; 48pp; English.

CC This is the amino acid sequence of a novel murine SMAD interacting
CC protein, designated S1P1. It is deduced from the sequence of a
CC cDNA clone (see AAV72502) isolated from mouse embryo cDNA using a
CC yeast two-hybrid screen. S1P1 represents a novel member of a
CC family of DNA-binding proteins including vertebrate delta-crystallin
CC enhancer binding proteins and Drosophila zfh-1. The C-terminal
CC zinc finger cluster of S1P1 (S1P1zf1) binds to E2 box sites and to
CC the Brachyury protein binding site (see AAV72511), and interferes
CC with Brachyury-mediated transcription activation in cells. It also
CC interacts with the C-domain of Smad 1, 2 and 5. It may be involved
CC in transcriptional regulation of important differentiation genes in
CC significant biological processes such as cell growth and
CC differentiation, embryogenesis, and abnormal cell growth, including
CC cancer. The invention also provides recombinant expression vectors,
CC transfects or transduced cells, a method of screening for compounds
CC which affect the interaction between SMAD and SMAD interacting
CC protein, transgenic animals useful for testing medicaments and as
CC therapy models, and a method for post-transcriptional regulation of
CC gene expression by members of the TGF-beta superfamily by
CC manipulation or modulation of the interaction between Smad function
CC and/or activity and mRNA stability. S1P1s and nucleic acids encoding
CC them are useful therapeutically and in claimed kits for diagnosing
CC diseases or disorders related to cancer, malformation, immune or
CC neural disease, bone metabolism related diseases or disorders and
CC diseases affecting organs such as skin, lung, kidney, pancreas,
CC stomach, gonad, muscle or intestine.
CC (Updated on 20-MAR-2003 to correct PA field.)

CC Sequence 944 AA;

Query Match 100.0%; Score 4991; DB 20; Length 944;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 944; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTTGAGNKKFKTECGAKFYKYLKHEHLRHSGEKPECPNKKRFSHSGSYSHISS 60
DB 1 MTTGAGNKKFKTECGAKFYKYLKHEHLRHSGEKPECPNKKRFSHSGSYSHISS 60
QY 61 KKCIGLISVNGRMNNIKTGSSPNVSSTPTNSAITQLRNKLENGKPLSMSEQTGLIKIT 120
DB 61 KKCIGLISVNGRMNNIKTGSSPNVSSTPTNSAITQLRNKLENGKPLSMSEQTGLIKIT 120
QY 121 TEPILDNDYKVLMTATGFGSSPFNNCGIATSPILGVHSAOSPMOHLGVGMARPLLGGP 180
DB 121 TEPILDNDYKVLMTATGFGSSPFNNCGIATSPILGVHSAOSPMOHLGVGMARPLLGGP 180
QY 181 TMSNLSSEYQVLYQVNDTVSRQMDCKTEDISKLYNHNKDCSQPBEQGVTSPIPIPV 240
DB 181 TMSNLSSEYQVLYQVNDTVSRQMDCKTEDISKLYNHNKDCSQPBEQGVTSPIPIPV 240
QY 241 GLPVVSHNGATSIIDYTLKVNKAACIQLSTTDSRROISNIKKERLTLIDLVTDK 300
DB 241 GLPVVSHNGATSIIDYTLKVNKAACIQLSTTDSRROISNIKKERLTLIDLVTDK 300
QY 301 TENHSTSTPSCQFCESPPGPPILOHERYLCKNNEIKAVIQPHENIVPNAGVFN 360
DB 301 TENHSTSTPSCQFCESPPGPPILOHERYLCKNNEIKAVIQPHENIVPNAGVFN 360
QY 361 KALLLSVSEKGLTSPINPYKDHMSVLYKAYYAMNPNDSDELKTSIVAGLPOEFVKEM 420
DB 361 KALLLSVSEKGLTSPINPYKDHMSVLYKAYYAMNPNDSDELKTSIVAGLPOEFVKEM 420
QY 421 FEQRKVVQYSNRSRPSLERTSKPLAPNSNPTTKDSLPRSPVYKPMDSITSPIAELHNSV 480

DB 421 FEQRKVVQYSNRSRPSLERTSKPLAPNSNPTTKDSLPRSPVYKPMDSITSPIAELHNSV 480
QY 481 TSCDPLRLTKSSHFNINIAVDKLDHSRNTSPNLNLSSTSSKNSHSSSTPSPSSSEL 540
DB 481 TSCDPLRLTKSSHFNINIAVDKLDHSRNTSPNLNLSSTSSKNSHSSSTPSPSSSEL 540
QY 541 QAEPLDLSLPGKQREBKGIATKNTKATSNLDHNSVSSSENSEPLNTPFIKKEFSN 600
DB 541 QAEPLDLSLPGKQREBKGIATKNTKATSNLDHNSVSSSENSEPLNTPFIKKEFSN 600
QY 601 SNLDMKSNPNVGMNPFSAKPLYPPLPOSAFPPTMPVQTSIPGLRPYGLDQMSF 660
DB 601 SNLDMKSNPNVGMNPFSAKPLYPPLPOSAFPPTMPVQTSIPGLRPYGLDQMSF 660
QY 661 LPHMAYTPTGATFADMOQRKRYORQKQFQGLDGAQDYNGLDMDTDSCLSRKTI 720
DB 661 LPHMAYTPTGATFADMOQRKRYORQKQFQGLDGAQDYNGLDMDTDSCLSRKTI 720
QY 721 KTTESGMYACDLCDKTFQKSSSLRHKYEHNTGRPHQCICCKAFKGLHLEHSLHSG 780
DB 721 KTTESGMYACDLCDKTFQKSSSLRHKYEHNTGRPHQCICCKAFKGLHLEHSLHSG 780
QY 781 EKPYQCDKCGKXPSHSGSYSOHNNHRYCYCKEAEEREAERARKKGLGPTLELMNRA 840
DB 781 EKPYQCDKCGKXPSHSGSYSOHNNHRYCYCKEAEEREAERARKKGLGPTLELMNRA 840
QY 841 YQSTIPQGSYSEERESMPROGESEKHEKGESEGYGLRRDGESESESESESESE 900
DB 841 YQSTIPQGSYSEERESMPROGESEKHEKGESEGYGLRRDGESESESESESESE 900
QY 901 MDTPETIRDEBETGDSMDSDGKMETKSDHEEDNMDGNG 944
DB 901 MDTPETIRDEBETGDSMDSDGKMETKSDHEEDNMDGNG 944

RESULT 2

ABB21853
ID ABB21853 standard; Protein; 533 AA.

AC ABB21853;

DT 23-JAN-2002 (first entry)

DE Protein #3852 encoded by probe for measuring heart cell gene expression.

KW Human; gene expression; heart; microarray; vascular system;

KW cardiovascular disease; hypertension; cardiac arrhythmia;

OS Homo sapiens.

PN MO200157274-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WC-US00666.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI, 2001-488899/53.

PT Single exon nucleic acid probes for analyzing gene expression in human

PT hearts -
XX
XX Claim 15; SEQ ID No 23623; 530bp; English.
PS
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart (see
CC AB21535-ABA41305). The present sequence is a protein encoded by one such
CC probe. The probe may be used for predicting, measuring and displaying
CC gene expression in samples derived from the human heart via microarrays.
CC By measuring gene expression, the probes are useful for predicting,
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPRO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 533 AA:
SQ
Query Match 52.3%; Score 2612; DB 22; Length 533;
Best Local Similarity 94.2%; Pred. No. 5.3e-177;
Matches 502; Conservative 12; Mismatches 19; Indels 0; Gaps 0;
QY 36 EKPYPENCKKRFHSHSGSYSSHSKCKICIGISVNGMRNNIKTGSSPNSVSSSPTNSAI 95
DB 1 EKPYPENCKKRFHSHSGSYSSHSKCKICIGISVNGMRNNIKTGSSPNSVSSSPTNSAI 60
QY 96 TOLRNKLENGKPLMSSEOTGLKIKTEPLDFNDYKVLMAHGFSGSSPNNGLGATSP 155
DB 61 TOLRNKLENGKPLMSSEOTGLKIKTEPLDFNDYKVLMAHGFSGSSPNNGLGATSP 120
QY 156 GVHPSAOSPMOHLGVGMEAPLLGFPPTNNSLSEVOKVLOIVDNTVSRQMDCKTEDISK 215
DB 121 GVHPSAOSPMOHLGVGMEAPLLGFPPTNNSLSEVOKVLOIVDNTVSRQMDCKTEDISK 180
QY 216 KGYHMKDPCSQPEBQGVTSPIPVGLPVVSHNGATKSIIDYTEKNEAACAQSLTTD 275
DB 181 KGYHMKDPCSQPEBQGVTSPIPVGLPVVSHNGATKSIIDYTEKNEAACAQSLTTD 240
QY 276 SRROISNKKKKRKLTLIDLVTDDKMIENHNSISTPSCQFCESPPGPIPLHOHERYLC 335
DB 241 SRROISNKKKKRKLTLIDLVTDDKMIENHNSISTPSCQFCESPPGPIPLHOHERYLC 300
QY 336 NEETKAVLOPHENIVPNKAGVFNKALLSSVSEKGLTSPINPYKDHMSVLAAYYAMN 395
DB 301 NEETKAVLOPHENIVPNKAGVFNKALLSSVSEKGLTSPINPYKDHMSVLAAYYAMN 360
QY 396 MEPSDELKISIAVGLPOEFVKEMFEQRKYQYNSNSPSLERTSKPLAPNSPTTKDS 455
DB 361 MEPSDELKISIAVGLPOEFVKEMFEQRKYQYNSNSPSLERTSKPLAPNSPTTKDS 420
QY 456 LLPRSPVKPMDSTSPSIAELHNSVTSCDPLRLTKSSHFTNITKAVDLHRSRNTSP 515
DB 421 LLPRSPVKPMDSTSPSIAELHNSVTSCDPLRLTKSSHFTNITKAVDLHRSRNTSP 480
QY 516 NLSTSSKNSHSSSYTPNSFSSEELQAEPLDLSPKORREPKGIATKNTKA 568
DB 481 NLSTSSKNSHSSSYTPNSFSSEELQAEPLDLSPKORREPKGIATKNTKA 533
RESULT 3
ID AAM57277 standard; Protein; 533 AA.
XX AAM57277;
AC
XX
XX 05-NOV-2001 (first entry)
DT
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 29382.
XX
XX Human brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
epilepsy; cancer.

XX
OS Homo sapiens.
XX
XX
XX WO200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
XX Example 4; SEQ ID NO: 29382; 650bp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention.
XX
XX Sequence 533 AA;
SQ
Query Match 52.3%; Score 2612; DB 22; Length 533;
Best Local Similarity 94.2%; Pred. No. 5.3e-177;
Matches 502; Conservative 12; Mismatches 19; Indels 0; Gaps 0;
QY 36 EKPYPENCKKRFHSHSGSYSSHSKCKICIGISVNGMRNNIKTGSSPNSVSSSPTNSAI 95
DB 1 EKPYPENCKKRFHSHSGSYSSHSKCKICIGISVNGMRNNIKTGSSPNSVSSSPTNSAI 60
QY 96 TOLRNKLENGKPLMSSEOTGLKIKTEPLDFNDYKVLMAHGFSGSSPNNGLGATSP 155
DB 61 TOLRNKLENGKPLMSSEOTGLKIKTEPLDFNDYKVLMAHGFSGSSPNNGLGATSP 120
QY 156 GVHPSAOSPMOHLGVGMEAPLLGFPPTNNSLSEVOKVLOIVDNTVSRQMDCKTEDISK 215
DB 121 GVHPSAOSPMOHLGVGMEAPLLGFPPTNNSLSEVOKVLOIVDNTVSRQMDCKTEDISK 180
QY 216 KGYHMKDPCSQPEBQGVTSPIPVGLPVVSHNGATKSIIDYTEKNEAACAQSLTTD 275
DB 181 KGYHMKDPCSQPEBQGVTSPIPVGLPVVSHNGATKSIIDYTEKNEAACAQSLTTD 240
QY 276 SRROISNKKKKRKLTLIDLVTDDKMIENHNSISTPSCQFCESPPGPIPLHOHERYLC 335
DB 241 SRROISNKKKKRKLTLIDLVTDDKMIENHNSISTPSCQFCESPPGPIPLHOHERYLC 300
QY 336 NEETKAVLOPHENIVPNKAGVFNKALLSSVSEKGLTSPINPYKDHMSVLAAYYAMN 395
DB 301 NEETKAVLOPHENIVPNKAGVFNKALLSSVSEKGLTSPINPYKDHMSVLAAYYAMN 360
QY 396 MEPSDELKISIAVGLPOEFVKEMFEQRKYQYNSNSPSLERTSKPLAPNSPTTKDS 455
DB 361 MEPSDELKISIAVGLPOEFVKEMFEQRKYQYNSNSPSLERTSKPLAPNSPTTKDS 420
QY 456 LLPRSPVKPMDSTSPSIAELHNSVTSCDPLRLTKSSHFTNITKAVDLHRSRNTSP 515
DB 421 LLPRSPVKPMDSTSPSIAELHNSVTSCDPLRLTKSSHFTNITKAVDLHRSRNTSP 480

	QY	516	NLSSTSSKNSHSSSYTPNPSFSEELQAEPDLISLPKORREPKGIATNKTKA	568
Dd		481	NLSSTSKSHSSSYYTPNPSFSFEELQAEPFLVITIKTERGKYXSHKEOKKS	533
	RESULT 4			
	ID	AAM69682	standard; Protein; 533 AA.	
	AC	AAM69682;		
	DT	06-NOV-2001	(first entry)	
	DE	Human bone marrow expressed probe encoded protein SEQ ID NO: 29988.		
	KM	Human; bone marrow expressed exon; gene expression analysis; probe; microarray; cancer; Leukaemia; Lymphoma; myeloma.		
	OS	Homo sapiens.		
	PW	MOL00157276-A2.		
	XD	09-AUG-2001.		
	PE	30-JAN-2001; 2001MO-US00668.		
	PR	04-FEB-2000; 2000US-0180312. 26-MAY-2000; 2000US-020745C. 30-JUN-2000; 2000US-060840B. 03-AUG-2000; 2000US-0632366. 21-SEP-2000; 2000US-0234687. 27-SEP-2000; 2000US-0236359. 04-OCT-2000; 2000GB-0024263.		
	PA	(MOLE-) MOLECULAR DYNAMICS INC.		
	PI	Penn SG, Hanzel DK, Chen W, Rank DR;		
	XX	WI; 2001-488900/53.		
	PT	Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human bone marrow -		
	PS	Example 4; SEQ ID NO: 29988; 658bp + Sequence Listing; English.		
	CC	The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human bone marrow. They can be used to measure gene expression in bone marrow samples, which may enable the improved diagnosis and treatment of cancers such as lymphoma, leukaemia and myeloma. The present sequence is a protein encoded by one of the probes of the invention.		
	SO	Sequence 533 AA:		
	Query Match	52.3%; Score 2612; DB 22; Length 533;		
	Best Local Similarity	94.2%; Pred. No. 5.3e-177;		
	Matches 502; Conservative	12; Mismatches 19; Indels 0; Gaps 0,		
QY		36 EKPYECPCKRRPSSHSGSYSHISKKCIGLIVNGRNRNNKITGSPPNVSSSPTNNAI	95	
Dd		1 EKPYPCKRKRPSSHSGSYSHISKKGILSVNGRMNRNNKITGSSPNVSASSPTNSAI	60	
QY		TQLRNKLGNPKPLMSSEOTGLLKITEPLDPNDYKVLAHTGFSGSPFMNGLGATSPL	155	
Dd		61 TQLRNKLGNKRPLMSSEOTGLKKITERPDNDYKVLAAHTGFSGTSPMNGLGATSBL	120	
QY		GVPBSAGSPMHLDGVMEAPLLGGFTPMNSNLSEVGKIQVDNTVSRQMDCKTEDISKTL	215	
Dd		121 GVHPBASGPMDHLGVMEAPLLGGFTPMNSNLSEVGKIQVDNTVSRQMDCKAETSKTL	180	
QY		KGYHHMKDSCSQPEEGVTSPNPRLPGLVYSANGATKSIDLTLEKNFAKCLDSLTDD	275	

Db	18	KGYHMKDPCSGDEQGVTSPIINPGLPVSHNGATKSIIIDYLEKVNEMAKCIQSLTTD	240
Qy	276	SRRLQISNIKKELKRLTLDLVYDDKKIENHISITPSCQFCESPPGPIPLHQRHYLCM	3353
Db	241	SRRLQISNIKKELKRLTLDLVYDDKKIENHISITPSCQFCESPPGPIPLHQRHYLCM	3000
Qy	336	NERIKAVLQPHENIYFNKAGVFDVKALLSSVSEKGLTSPINPYQDHMSVLKAYYAMN	3959
Db	301	NERIKAVLQPHENIYFNKAGVFDVKALLSSVSEKGLTSPINPYQDHMSVLKAYYAMN	3660
Qy	396	MEBNSDELLKISIAVGLPQEFVKWFEQRYQYQNSRSPSLERTSKPLAPNSNPTKDS	4555
Db	361	MEBNSDELLKISIAVGLPQEFVKWFEQRYQYQNSRSPSLERTSKPLAPNSNPTKDS	4200
Qy	456	LPLRSYVKKQMSDITSPSIAELHNSVTCDDPPLRLTKSHFNINIAVDKLDHSRNTSPL	5151
Db	421	LPLRSYVKKQMSDITSPSIAELHNSVTCDDPPLRLTKSHFNINIKVPEKLDHSRNTSPL	4800
Qy	516	NLSSTSKSHSSSYTPNSFSSEELQAPLPLSLPKQWRBPKGIATKNTKA	568
Db	481	NLSSTSKSHSSSYTPNSFSSEELQAPLPLSLPKQWRBPKGIATKNTKA	533
RESULT 5			
ID	AAM30015		
AC	AAM30015	standard; Protein; 533 AA.	
DT	17-OCT-2001	(first entry)	
DE	Peptide #4052 encoded by probe for measuring placental gene expression.		
KM	Probe; microarray; human; placenta; antenatal diagnosis;		
OS	genetic disorder.		
XX	Homo sapiens.		
XX	WO200157272-A2.		
PD	09-AUG-2001.		
XX	30-JAN-2001; 2001MO-US00663.		
XX	04-FEB-2000; 2000US-0180312.		
PR	26-MAY-2000; 2000US-0207456.		
PR	30-JUN-2000; 2000US-0608408.		
PR	03-AUG-2000; 2000US-0632366.		
PR	21-SEP-2000; 2000US-0234687.		
PR	27-SEP-2000; 2000US-0236359.		
PR	04-OCT-2000; 2000GB-0024263.		
XX	(MOLB-) MOLECULAR DYNAMICS INC.		
XX	Penn SG, Hanzel DK, Chen W, Rank DR;		
XX	WPI; 2001-488897/53.		
PT	Human genome-derived single exon nucleic acid probes useful for		
PT	analyzing gene expression in human placenta -		
XX	Claim 27; SEQ ID No 30284; 654bp; English.		
CC	The present invention relates to single exon nucleic acid probes (SENP;		
CC	see A113315-A157546). The present sequence is a peptide encoded by one		
CC	such probe. The probes are useful for producing a microarray for		
CC	predicting, measuring and displaying gene expression in samples derived		
CC	from human placenta. The probes are useful for antenatal diagnosis of		
CC	human genetic disorders.		
XX	Sequence	533 AA;	
XX	Query Match	52.3%; Score 2612; DB 22; Length 533;	

Best Local Similarity 94.2%, Pred. No. 5.3e-177;
Matches 502; Conservative 12; Mismatches 19; Indels 0; Gaps 0;

```

QY 36 EKPYECPCCKRPFHSGSYSHISKKCTIGLISVNGRMNNIKTGSSPNSVSSPTNSAI 95
   |||||
DB 1 EKPYECPCCKRPFHSGSYSHISKKCTIGLISVNGRMNNIKTGSSPNSVSSPTNSAI 60
   |||||

QY 96 TOLANKLENGKPLSMSEQTGLLKI KTEPLDPNDYKVLMTATGFGSSPPMNGIGATSP 155
   |||||
DB 61 TOLANKLENGKPLSMSEQTGLLKI KTEPLDPNDYKVLMTATGFGSGTSPFMNGIGATSP 120
   |||||

QY 156 GVHPSAOSPMDHGVGMEAPLGLPPTMNSLSEVOKLQIVDNTVSRKMDCKEDISK 215
   |||||
DB 121 GVHPSAOSPMDHGVGMEAPLGLPPTMNSLSEVOKLQIVDNTVSRKMDCKEDISK 180
   |||||

QY 216 KGYHMKDCSQPEEGQVTSPIPIPGVPSHNGATKSIIDYTLKVEAKACIQSLT 275
   |||||
DB 181 KGYHMKDCSQPEEGQVTSPIPIPGVPSHNGATKSIIDYTLKVEAKACIQSLT 240
   |||||

QY 276 SRROIINIKKEKRLTLDLVTDKMIENHSISTPSCQCKESPPGPPLHQBHERYLCKM 335
   |||||
DB 241 SRROIINIKKEKRLTLDLVTDKMIENHSISTPSCQCKESPPGPPLHQBHERYLCKM 300
   |||||

QY 336 NEETKAVIQPHENIVPNKAGVFDNKALLSSVLSSEKGLTSPINPYKDHMSVLYKAYTAN 395
   |||||
DB 301 NEETKAVIQPHENIVPNKAGVFDNKALLSSVLSSEKGMTSPINPYKDHMSVLYKAYTAN 360
   |||||

QY 396 MEPSDELKTSIANGLPQEPFKMEFEQRKAYOYNSNSPSPSLERSKPLANSNPPTKDS 455
   |||||
DB 361 MEPSDELKTSIANGLPQEPFKMEFEQRKAYOYNSNSPSPSLERSKPLANSNPPTKDS 420
   |||||

QY 456 LLPSRPVPMDSITSPSIAELHNSVTSQDPLRLTKSSHFTNIKAVDKLDSRSNTSP 515
   |||||
DB 421 LLPSRPVPMDSITSPSIAELHNSVTSQDPLRLTKSSHFTNIKAVDKLDSRSNTSP 480
   |||||

QY 516 NLSTSSKNSHSSSTTTPNSFSSEELQAPPLDLSPKQREPKGIATNTKTA 568
   |||||
DB 481 NLSTSSKNSHSSSTTTPNSFSSEELQAPPLDLSPKQREPKGIATNTKTA 533
   |||||

RESULT 6
ABG39298
ID ABG39298 standard; Peptide; 533 AA.
XX
AC ABG39298;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human peptide encoded by genome-derived single exon probe SEQ ID 28963.
XX
KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karsagen syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease.
XX
OS Homo sapiens.
XX
PN WO200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US00665.
XX
PR 04-FEB-2000; 2000US-180312P.
PR 26-MAY-2000; 2000US-207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-234687P.

```

PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2002-114183/15.
XX
PT Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples -
XX
PS Claim 27; SEQ ID No 28963; 634pp; English.

The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human lung comprising single exon nucleic acid probes having one of 12614 nucleic acid sequences mentioned in the specification, or their complements or the 12387 open reading frames derived from the 12614 probes. Also included are a microarray comprising the novel set of probes; the novel set of probes which hybridize at high stringency to a nucleic acid expressed in the human lung; measuring gene expression in a sample derived from human lung, comprising (a) contacting the array with a collection of detectably labeled nucleic acids derived from human lung mRNA, and (b) measuring the label detectably bound to each probe of the array; identifying exons in a eukaryotic genome, comprising (a) algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karsagen syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a peptide/protein encoded by a single exon probe of the invention.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 533 AA;

Query Match 52.3%; Score 2612; DB 23; Length 533;
Best Local Similarity 94.2%; Pred. No. 5.3e-177;
Matches 502; Conservative 12; Mismatches 19; Indels 0; Gaps 0;

```

QY 36 EKPYECPCCKRPFHSGSYSHISKKCTIGLISVNGRMNNIKTGSSPNSVSSPTNSAI 95
   |||||
DB 1 EKPYECPCCKRPFHSGSYSHISKKCTIGLISVNGRMNNIKTGSSPNSVSSPTNSAI 60
   |||||

QY 96 TOLANKLENGKPLSMSEQTGLLKI KTEPLDPNDYKVLMTATGFGSSPPMNGIGATSP 155
   |||||
DB 61 TOLANKLENGKPLSMSEQTGLLKI KTEPLDPNDYKVLMTATGFGSGTSPFMNGIGATSP 120
   |||||

QY 156 GVHPSAOSPMDHGVGMEAPLGLPPTMNSLSEVOKLQIVDNTVSRKMDCKEDISK 215
   |||||
DB 121 GVHPSAOSPMDHGVGMEAPLGLPPTMNSLSEVOKLQIVDNTVSRKMDCKEDISK 180
   |||||

```

QY 216 KGYHMKDPCSGPEQGVTSNPINPVGLPVVSHNGATKSIIDYTLKYNKAKCLOSLTID 275
 Db 181 KGYHMKDPCSGPEQGVTSNPINPVGLPVVSHNGATKSIIDYTLKYNKAKCLOSLTID 240
 QY 276 SRROINNIKEKRTTIDLTVDKMIENHSISTPSCQFCKESPPGPIPLHGERLYCKK 335
 Db 241 SRROINNIKEKRTTIDLTVDKMIENHSISTPSCQFCKESPPGPIPLHGERLYCKK 300
 QY 336 NEEIKAVLOPHENIVPNKACVPVNDKALLISVLSSEKGLTSPINPYKDHMSVLYKAYVAMN 395
 Db 301 NEEIKAVLOPHENIVPNKACVPVNDKALLISVLSSEKGLTSPINPYKDHMSVLYKAYVAMN 360
 QY 396 MEENSELKISTAVGLPOEFVKEWEPQKRVQYNSRSPSLERTSKPLAPNSNPTTKS 455
 Db 361 MEENSELKISTAVGLPOEFVKEWEPQKRVQYNSRSPSLERTSKPLAPNSNPTTKS 420
 QY 456 LLEPRSVKPMDSITSPSIAELHNSVTSCEPPLRLTSSHFTNIKAVDKLDHSRNTSPL 515
 Db 421 LLEPRSVKPMDSITSPSIAELHNSVTSCEPPLRLTSSHFTNIKAVDKLDHSRNTSPL 480
 QY 516 NUSSTSSKNSHSSSYTPNSFSSEELQAEPLDLSLPKQMRBPKGIATKNTKA 568
 Db 481 NUSSTSSKNSHSSSYTPNSFSSEELQAEPLDLSLPKQMRBPKGIATKNTKA 533
 RESULT 7
 ABUS7636
 ID ABUS7636 standard; Protein; 1125 AA.
 AC ABUS7636;
 XX
 DT 09-APR-2003 (first entry)
 DE Differentially expressed breast cancer associated protein #23.
 XX
 KM Breast cancer; differential gene expression; BC-cDNA;
 KM breast cancer diagnosis; breast cancer monitoring;
 KM breast cancer treatment; breast cancer staging.
 XX
 OS Homo sapiens.
 XX
 PN US2002156263-A1.
 XX
 PD 24-OCT-2002.
 XX
 PF 04-OCT-2001; 2001US-0974298.
 XX
 PR 05-OCT-2000; 2000US-238331P.
 XX
 PA (CHEN/) CHEN H.
 XX
 PI Chen H;
 XX
 DR WPI; 2003-182653/18.
 XX
 PT New cDNAs, which are differentially expressed in (metastatic) breast
 PT cancer useful for diagnosing or staging, breast cancer, or for
 PT monitoring the treatment of breast cancer in an individual -
 XX
 PS Example; SEQ ID NO 114; 30pp; English.
 XX
 CC The invention describes a combination of cDNAs (designated BC-cDNAs),
 CC which are differentially expressed in breast cancer. The combination
 CC includes 152 cDNA sequences, or their complements. The protein encoded
 CC by any of these BC-cDNAs is useful for screening several molecules or
 CC compounds to identify at least one ligand that specifically binds the
 CC protein, producing or preparing polyclonal or monoclonal antibodies, or
 CC purifying antibodies from a sample. The antibodies, which specifically
 CC bind the protein differentially expressed in breast cancer is useful for
 CC detecting the expression of a protein in a sample. The BC-cDNAs are
 CC also useful for diagnosing, monitoring the treatment of, or staging,
 CC breast cancer. This is the amino acid sequence of a differentially
 CC expressed breast cancer associated protein.

CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from the US patent office at
 CC seqdata.uspto.gov/sequence.html?docid=20020156263.

SQ Sequence 1125 AA;

Query Match 33.2%; Score 1657.5; DB 24; Length 1125;

Best Local Similarity 41.9%; Pred. No. 1.2e-108;

Matches 406; Conservative 147; Mismatches 308; Indels 107; Gaps 32;

QY 2 LFGAGNRKFKCEGKAFKYGKHLKHLRHSGEKPYCPNCKRFSHSGSYSHISRK 61
 Db 232 VTQSGCNRKFKCEGKAFKYGKHLKHLRHSGEKPYCPNCKRFSHSGSYSHISRK 291
 QY 62 KICGLISVNGRRNNIKTG--SSPNSVSSPTNSAITGLRNLKNGKPLMSBQGLTKI 119
 Db 292 KICSLIPVNGRRRTGLTKTQCSPP--SLSPSPSPRRPQKRTIEN--QGLSVNGI 347
 QY 120 KTEPLDFNDYKVLMTATGSGSSPPFANGGLATSPGVNPSAQSPYQHL---GYGAEPL 176
 Db 348 KTEPLVDY--EFKPIVVASGNGCSTPLONGVFTGGPQATSPQGWAVVLPVGLVSP 406
 QY 177 LGPPTNNSLSEYQKVLQI--VDNTVSRQKMDCKTEDISKLGKGYHMKDPCSGPEQGVTS 235
 Db 407 -----SINLSDIQNVKAVDGNVIRQVLE--NNQANLAKGEQETINASPIQCGSHV 457
 QY 236 NIPVGLPVVSHNGATKSIIDYTLKYNKAKCLOSLTIDSRROINNIKEKRTTIDLV 295
 Db 458 -ISAISLPLVDQGTTKIITINSLBQPSQLQVPPQLKKGNPVATVSCSEKLPEDLYVK 516
 QY 296 TD-DKMIENH-SISTPSCQFCKESPPGPI-PLHGERLYCKMNEIKAVLOPHENIVPN 352
 Db 517 SEKDKSPEGGVNDSTCLDCCD---PGDINALPELKHV-----DLKQPTP---BP 561
 QY 353 KAGVFDVNDKALLISVLSSEKGLTSPINPYKDHMSVLYKAYVAMNMEENSELKISTAVGL 412
 Db 562 LPAELAEKRESSVSAITGNGNLSPSOPPLKNLSLKAAYALNAQSAEELSIALDSVNL 621
 QY 413 POEFVKEWEPQKRVQYNSRSPSLERTSKPLAPNSNPTTKOSLLEPRSVKPMDSITSPS 472
 Db 622 PLDVVVKWEPKQAGIS-----VQSESPSPBERKV-----NTPAKNNDDQQSAN 667
 QY 473 IAEIANSVTSCEPPLRLTSSHFTNIKAVDKLDHSRNTSPLNLSSTSSKNSHSSSYTP 532
 Db 668 ANEPQDSTVNLQSLPKMTNS---PVLPGVSTNGSSSTSPSPPLNLSSTRTQGYLYTA 724
 QY 533 NSPFSSEELQAEPLDLSLPKQMRBPKGIATKNTKATSIINLHNSVSSSENS-----DE 587
 Db 725 EG-AQEEPOVEPLDLSLPKQGE-----LLERSTTIVYQNSVYVQGE 767
 QY 588 PLNLTFIKKEFSNNSNLDNKSNPVFGMNPFSAPLYTPLPQSAFPATFMPVQTSIP 647
 Db 768 PLNLSCAKKEPQDSCVTD--SEPVVNVIPPSANPINIAIPVTAQLPTVIAIDQNSVP 825
 QY 648 GLRPTFGDLQMSLPLMAATYTPGATFPADMQQRKRYQKQGFQGLLDGADYMSGLD 707
 Db 826 CRLALANQKTLIPQVAYTSTVSP-AVQEPPLKLVIPNGQDERQDTSSBGSVND 884
 QY 708 MTPDSCLSRKKIKKTESGVYACDLCTFKFQSSSLRLRKHETGRPQCOCKKAFNG 767
 Db 885 QNDSIDSTPPKKKARKTENGMVACDLCDKIFQSSSLRLRKHETGRPHECGCKKAFNG 944
 QY 768 KHLILHSHRLHSGEKPYQDCGKRFPSHSGSYQHNHRYSYCKRBAEREAERAREK 827
 Db 945 KHLILHSHRLHSGEKPYQDCGKRFPSHSGSYQHNHRYSYCKRBAERERDSTGEQ-----1000
 QY 828 GHIGPTLEMLNRAVLOSTIP-QGYSSERESMPR--DESEKENEKGEF-----GYK 879
 Db 1001 -EAGP--EILSNHNVGARAPSQG--DSDERESTIREEDDSKEEEDKEMELQOEKE 1056
 QY 880 LRRDDEDEEEEEESEKSMQTDPEITRDEEETDHSMDSSSD-----GKMET 930

Db 1057 CEKPGDEEEEEEVEEVE-BAENGBEAKTEGLMKODRAESQASLSGQKVESSE 1115

QY 931 KSDHEEDN 938

Db 1116 QVSEKTN 1123

RESULT 8

ABG48956

ID ABG48956 standard; Peptide; 190 AA.

XX

AC ABG48956;

XX

DT 25-FEB-2003 (first entry)

XX

DE Human liver peptide, SEQ ID No 27604.

XX

KM Human, liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;

KM hypercholesterolaemia; coronary heart disease.

XX

OS Homo sapiens.

XX

PN WO200157273-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00664.

XX

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488898/53.

XX

PT Human genome-derived single exon nucleic acid probes useful for

PS analysing gene expression in human adult liver -

XX

PS Claim 27; SEQ ID No 27604; 658bp; English.

XX

CC The invention relates to a single exon nucleic acid probe (SENP) (1) for

CC measuring human gene expression in a sample derived from human adult

CC liver, comprising one of 13109 defined nucleotide sequences given in the

CC specification (or complements/ fragments). The probe hybridises at high

CC stringency to a nucleic acid molecule expressed in the human adult

CC liver. (1) may be used for predicting, measuring and displaying gene

CC expression in samples derived from human adult liver. The genes

CC identified may be involved in genetic liver diseases such as cirrhosis,

CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which

CC is associated with coronary heart disease. ABG47348-ABG5930 represent

CC human liver single exon encoded peptides of the invention.

CC Note: The sequence information for this patent does not appear in the

CC printed specification but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX

XX Sequence 190 AA;

XX

Query Match 20.1%; Score 1005; DB 22; Length 190;

Best Local Similarity 97.4%; Pred. No. 2e-63;

Matches 185; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 753 KRPHOCQCCKAFKFKHHLIHSRLHSGEKPYQCCGKRFSGSHSGOHMHRYSYCKR 812

Db 1 KRPHOCQCCKAFKFKHHLIHSRLHSGEKPYQCCGKRFSGSHSGOHMHRYSYCKR 60

QY 813 EAEREAERAREKGLPTLIMNRAVLOSITPGYSDSEERSEMPRDSSEKHEKE 872

Db 61 EAEREAERAREKGLPTLIMNRAVLOSITPGYSDSEERSEMPRDSSEKHEKE 120

QY 873 GEQYGLKRRDGDDEEESEENKSMOTDPTIRDEETGHSMDSSGDKMETKS 932

Db 121 GEDYGLKRRDGDDEEESEENKSMOTDPTIRDEETGHSMDSSGDKMETKS 180

QY 933 DHEEDNMDG 942

Db 181 DHEEDNMDG 190

RESULT 9

ABG51171

ID ABG51171 standard; Peptide; 190 AA.

XX

AC ABG51171;

XX

DT 25-FEB-2003 (first entry)

XX

DE Human liver peptide, SEQ ID No 29819.

XX

KM Human, liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;

KM hypercholesterolaemia; coronary heart disease.

XX

OS Homo sapiens.

XX

PN WO200157273-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00664.

XX

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488898/53.

XX

PT Human genome-derived single exon nucleic acid probes useful for

PS analysing gene expression in human adult liver -

XX

PS Claim 27; SEQ ID No 29819; 658bp; English.

XX

CC The invention relates to a single exon nucleic acid probe (SENP) (1) for

CC measuring human gene expression in a sample derived from human adult

CC liver, comprising one of 13109 defined nucleotide sequences given in the

CC specification (or complements/ fragments). The probe hybridises at high

CC stringency to a nucleic acid molecule expressed in the human adult

CC liver. (1) may be used for predicting, measuring and displaying gene

CC expression in samples derived from human adult liver. The genes

CC identified may be involved in genetic liver diseases such as cirrhosis,

CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which

CC is associated with coronary heart disease. ABG47348-ABG5930 represent

CC human liver single exon encoded peptides of the invention.

CC Note: The sequence information for this patent does not appear in the

CC printed specification but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX

XX Sequence 190 AA;

XX

Query Match 20.1%; Score 1005; DB 22; Length 190;

Best Local Similarity 97.4%; Pred. No. 2e-63;

Matches 185; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy	753	KRHHQCOICKKAFKHGHHIIHESRLHSGKPYQCCCKCRFSHSGSYSOHNNHYSYCKR	8112
Db	1	KRHQCOICKKAFKHGHHIIHESRLHSGKPYQCCCKCRFSHSGSYSOHNNHYSYCKR	60
Qy	813	EAEEREAERAREKCHLAPTELLMNRAYLOSITPOGYSDEERESBPRDGESEKEHEKE	8727
Db	61	EAEEREAERAREKCHLAPTELLMNRAYLOSITPOGYSDEERESBPRDGESEKEHEKE	120
Qy	873	GEGCYCKLARPRDDEEESESESNKSMDDPETIRPEESTGCHSMDDSEDCOKMTKS	9322
Db	121	GEGCYCKLARPRDDEEESESESNKSMDDPETIRPEESTGCHSMDDSEDCOKMTKS	180
Qy	933	DHEEDNMEDG 942	
Db	181	DHEEDNMEDG 190	
RESULT 10			
ID	ABB28955		
ID	ABB28955	strand; Peptide; 190 AA.	
AC	ABB28955;		
DT	01-FEB-2002	(first entry)	
XX			
DE	Peptide #1606 encoded by breast cell single exon nucleic acid probe.		
XX			
KW	Human; microarray; single exon probe; gene expression; breast;		
KW	disease; cancer.		
XX			
OS	Homo sapiens.		
XX			
PN	W0200157271-A2.		
XX			
PD	09-AUG-2001.		
XX			
PF	30-JAN-2001; 2001WO-US00662.		
PR	04-FEB-2000; 2000US-0180312.		
PR	26-MAY-2000; 2000US-0207456.		
PR	30-JUN-2000; 2000US-0608408.		
PR	03-AUG-2000; 2000US-0632366.		
PR	21-SEP-2000; 2000US-0234687.		
PR	27-SEP-2000; 2000US-0236359.		
PR	04-OCT-2000; 2000GB-0024263.		
XX			
PA	(MOLE-) MOLECULAR DYNAMICS INC.		
PI	Penn SG, Hanzel DK, Chen W, Rank DR;		
DR	WPI; 2001-496933/54.		
PT	New spatially-addressable set of single exon nucleic acid probes,		
PT	useful for measuring gene expression in sample derived from human		
PT	breast, comprises number of single exon nucleic acid probes -		
XX			
XX	Claim 27; SEQ ID NO 11923; 327bp + sequence listing; English.		
CC	The invention relates to a spatially-addressable set of single exon		
CC	nucleic acid probes for measuring gene expression in a sample derived		
CC	from human breast and BT 474 cells. The method involves contacting		
CC	the probes with a collection of detectably labeled nucleic acids		
CC	derived from mRNA of human breast, and then measuring the label		
CC	bound to each probe of the microarray. The probes are useful for		
CC	verifying the expression of regions of genomic DNA predicted to		
CC	encode proteins. They are useful for gene discovery, and for		
CC	determining predisposition and/or prognosing breast disease. Gene		
CC	expression analysis is useful for assessing the toxicity of chemical		
CC	agents on cells. The microarray of this invention presents a far greater		
CC	diversity of probes for measuring gene expression, with far less bias		
CC	than expressed sequence tag microarrays. The method is suitable for		
CC	rapid production of functional information from genomic sequence. The		
CC	present sequence is a peptide encoded by a single exon nucleic acid		

	CC	probe of the invention.
	CC	Note: The sequence data for this patent did not form part of the
	CC	printed specification, but was obtained in electronic format directly
	CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences.
	XX	
	SQ	Sequence 190 AA;
	Query Match	20.1%; Score 1005; DB 22; Length 190;
	Best Local Similarity	97.4%; Pred. No. 2e-63; Mismatches 3; Gaps 0;
	Matches 185; Conservative	2; Indels 0;
Oy	753	KRPHQCOICKAFYGHKHLLIEHSRLHSGEKPYCDCKGKRFSHGYSQHNNHRYSYCKR 812
Dd	1	KRPHQCOICKKAFGHKHLLIEHSRLHSGEKPYCDCKGKRFSHGYSQHNNHRYSYCKR 60
Oy	813	EAEEREAAREARAKRHGLGPPELLMNRAYLOSITPGYSDEEERSMPRDGESEKEHEKE 872
Dd	61	EAEEREAAREARAKRGHLPEPELLMNRAYLOSIPTGGYSDEEERSMPRDGESEKEHEKE 120
Oy	873	GEEGYGLRRDRGDGEEEEESSENSMDTPETIRPDEETIGHSMDDSSDGDGMETS 932
Dd	121	GEDDYGLGRDGDEEFEESESSENKSMDTPETIRDEETIGHSMDDSSDGDGMETS 180
Oy	933	DHEEDNMEDG 942
Dd	181	DHEEDNME DG 190
	RESULT 11	
	ABB31134	
	ABB31134 standard; peptide: 190 AA.	
XX	ABB31134;	
AC		
XX	01-FEB-2002 (first entry)	
DT		
XX		
DE	Peptide #3785 encoded by breast cell single exon nucleic acid probe.	
XX		
KW	Human; microarray; single exon probe; gene expression; breast;	
XX	disease; cancer.	
XX		
OS	Homo sapiens.	
XX		
FN	WO200157271-A2.	
XX		
PD	09-AUG-2001.	
XX		
PF	30-JAN-2001; 2001WO-US00662.	
XX		
PR	04-FEB-2000; 2000US-0180312.	
FR	26-MAY-2000; 2000US-0207456.	
PR	30-JUN-2000; 2000US-0608408.	
PR	03-AUG-2000; 2000US-0632366.	
PR	21-SEP-2000; 2000US-0234687.	
PR	27-SEP-2000; 2000US-0236359.	
PR	04-OCT-2000; 2000GB-0024263.	
XX		
PA	{MOLE-} MOLECULAR DYNAMICS INC.	
PI	Penn SG, Hanzel DK, Chen W, Rank DR;	
XX		
WPt	WIPI; 2001-496933/54.	
XX		
PT	New spatially-addressable set of single exon nucleic acid probes,	
FT	useful for measuring gene expression in sample derived from human	
PT	breast, comprises number of single exon nucleic acid probes -	
PS		
XX	Claim 27; SEQ ID NO 14102; 327bp + sequence listing; English.	
XX	The invention relates to a spatially-addressable set of single exon	
CC	nucleic acid probes for measuring gene expression in a sample derived	
CC	from human breast and BT 474 cells. The method involves contacting	
CC	the probes with a collection of detectably labelled nucleic acids	

PT analyzing gene expression in human fetal liver -
 PS Claim 27; SEQ ID NO 28968; 639pp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 190 AA;
 XX
 Query Match 20.1%; Score 1005; DB 22; Length 190;
 Best Local Similarity 97.4%; Pred. No. 2e-63;
 Matches 185; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 753 KRPHQCQICKKAFKXKHLIEHSRLHSGEKPYQCDKCGKRFSGSYSGHMHRYSYCKR 812
 Db 1 KRPHQCQICKKAFKXKHLIEHSRLHSGEKPYQCDKCGKRFSGSYSGHMHRYSYCKR 60
 QY 813 EAEREERAEAREKRGHLEPTLLMNRAYLQSTTPQGYSDSERESMPRDGSEKEHEKE 872
 Db 61 EAEREERAEAREKRGHLEPTLLMNRAYLQSTTPQGYSDSERESMPRDGSEKEHEKE 120.
 QY 873 GEEGYGKLRRDGDEEEEEESENKSMOTDPETIRDEETGDSMDSDSGDKMETKS 932
 Db 121 GEEGYGKLRQDGDDEFEFESEENKSMOTDPETIRDEETGDSMDSDSGDKMETKS 180
 QY 933 DHEEDNMEDG 942
 Db 181 DHEEDNMEDG 190
 RESULT 14
 ABB19564
 ID ABB19564 standard; Protein; 190 AA.
 XX
 AC ABB19564;
 DT 23-JAN-2002 (first entry)
 XX
 DE Protein #1563 encoded by probe for measuring heart cell gene expression.
 XX
 KM Human; gene expression; heart; microarray; vascular system;
 KM cardiovascular disease; hypertension; cardiac arrhythmia;
 KM congenital heart disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200157274-A2.
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00666.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI; 2001-488899/53.
 XX

PT Single exon nucleic acid probes for analyzing gene expression in human
 PS hearts -
 XX
 CC Claim 15; SEQ ID NO 21334; 530pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart (see
 CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
 CC probe. The probes may be used for predicting, measuring and displaying
 CC gene expression in samples derived from the human heart via microarrays.
 CC By measuring gene expression, the probes are useful for predicting,
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the
 CC human heart and vascular system e.g. cardiovascular disease,
 CC hypertension, cardiac arrhythmias and congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 190 AA;
 XX
 Query Match 20.1%; Score 1005; DB 22; Length 190;
 Best Local Similarity 97.4%; Pred. No. 2e-63;
 Matches 185; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 753 KRPHQCQICKKAFKXKHLIEHSRLHSGEKPYQCDKCGKRFSGSYSGHMHRYSYCKR 812
 Db 1 KRPHQCQICKKAFKXKHLIEHSRLHSGEKPYQCDKCGKRFSGSYSGHMHRYSYCKR 60
 QY 813 EAEREERAEAREKRGHLEPTLLMNRAYLQSTTPQGYSDSERESMPRDGSEKEHEKE 872
 Db 61 EAEREERAEAREKRGHLEPTLLMNRAYLQSTTPQGYSDSERESMPRDGSEKEHEKE 120
 QY 873 GEEGYGKLRRDGDEEEEEESENKSMOTDPETIRDEETGDSMDSDSGDKMETKS 932
 Db 121 GEEGYGKLRQDGDDEFEFESEENKSMOTDPETIRDEETGDSMDSDSGDKMETKS 180
 QY 933 DHEEDNMEDG 942
 Db 181 DHEEDNMEDG 190
 RESULT 15
 ABB21690
 ID ABB21690 standard; Protein; 190 AA.
 XX
 AC ABB21690;
 DT 23-JAN-2002 (first entry)
 XX
 DE Protein #3689 encoded by probe for measuring heart cell gene expression.
 XX
 KM Human; gene expression; heart; microarray; vascular system;
 KM cardiovascular disease; hypertension; cardiac arrhythmia;
 KM congenital heart disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200157274-A2.
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00666.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI; 2001-488899/53.
 XX

THIS PAGE BLANK (usr10)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2004, 16:27:45 ; Search time 28 Seconds
(without alignments)
3242.258 Million cell updates/sec

Title: US-09-964-238-2

Perfect score: 4991
Sequence: 1 MLTQAGNRKPKCTECGKAF.....DGKMETKSDHEBDNEDMG 944

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	4273	85.6	1214	2	JC7259 Smad interacting p
2	1665	33.4	1114	2	deltae1 - chicken
3	1657.5	33.2	1124	2	zinc finger protei
4	1605.5	32.2	1154	2	E-box-binding repr
5	1599	32.0	1043	2	DNA-binding protei
6	1587.5	31.8	1117	2	delta-crystallin/E
7	819.5	16.4	1060	2	homeotic protein z
8	741	14.8	680	2	hypothetical prote
9	379	7.6	1191	2	zinc finger protei
10	358	7.2	1350	2	finger protein - A
11	356.5	7.1	1042	2	Evil protein - hum
12	355.5	7.1	1051	2	oncogene Evi-1 - h
13	345	6.9	1042	2	transcription regu
14	328.5	6.6	701	2	hypothetical prote
15	315.5	6.3	803	2	zinc finger protei
16	312.5	6.1	469	2	zinc finger protei
17	302.5	6.1	1173	2	prockr2 - chicken
18	299.5	6.0	1615	2	ras-responsive ele
19	298.5	6.0	1186	2	O/E-1-associated z
20	296	5.9	580	2	specmatogenesis pr
21	296	5.9	675	2	zinc-finger protei
22	292	5.9	651	2	finger protein 2,
23	290.5	5.8	1350	2	finger finger protei
24	289.5	5.8	615	2	finger protein (cl
25	289.5	5.8	1323	2	spalt protein - mo
26	285.5	5.7	636	2	gene NK10 protein
27	285	5.7	460	2	hypothetical prote
28	285	5.7	856	2	B-lymphocyte-induc
29	284	5.7	710	2	zinc finger protei

30	280.5	5.6	511	2	150114	early growth respo
31	279	5.6	644	2	T46277	hypothetical prote
32	278.5	5.6	693	2	137570	zinc finger protei
33	277.5	5.6	536	2	S06548	finger protein (cl
34	276	5.5	553	2	S22954	finger protein zfp
35	275.5	5.5	706	2	A48752	B-cell CLL/Lymphom
36	274.5	5.5	508	2	A32225	nerve growth facto
37	274.5	5.5	789	2	A39564	transcription repr
38	273.5	5.5	727	2	B60191	transcription regu
39	272.5	5.5	543	2	A41211	early growth respo
40	271.5	5.4	706	2	152586	B-cell CLL/Lymphom
41	270	5.4	533	2	U50304	developmental cont
42	266	5.3	668	2	T08725	probable finger pr
43	265	5.3	733	2	JC7679	dendritic cell-der
44	264.5	5.3	678	2	S62939	hypothetical prote
45	264	5.3	707	2	S68858	finger protein - m

ALIGNMENTS

RESULT 1

Smad interacting protein 1 - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C:Accession: JC7259
R:Etaki, A.; Kuroda, H.; Fukui, A.; Aashima, M.
Biochem. Biophys. Res. Commun. 271, 151-157, 2000
A:Title: XSLP1, a member of two-handed zinc finger proteins, induced anterior neural mar
A:Reference number: JC7259
A:Accession: JC7259
A:Molecule type: mRNA
A:Residues: 1-1214 <EIS>
A:Experimental source: egg
C:Comment: This protein, a two-handed zinc finger and a homeodomain protein, is a transci
signal pathway. It plays the roles in early neural development.
C:Genetics:
A:Gene: xslp1
C:Keywords: egg; zinc finger

Query Match	85.6%	Score 4273	DB 2	Length 1214	Best Local Similarity 86.1%; Pred. No. 1.4e-215
Matches 815	Conservative 56	Mismatches 68	Indels 8	Gaps 5	
QY	1	MLTQAGNRKPKCTECGKAFKXKHLKEHLRHSGEKYECPCNCKKRSKSHSGVSSHTSS	60		
DB	272	MLTQAGNRKPKCTECGKAFKXKHLKEHLRHSGEKYECPCNCKKRSKSHSGVSSHTSS	331		
QY	61	KKICGLISVNGRMNNIKTGSSPNSVSSPTNSAITQLRNKLKNGKPLMSGEQTLKIK	120		
DB	332	KKICGLISVNGRMNNIKTGSSPNSVSSPTNSAITQLRNKLKNGKPLMSGEQTLKIK	391		
QY	121	TEPLDENDYKVM-ATHGPGSSGSPMNGLGATSPGLVHPSAQSPMGLGVMEAPLLGF	179		
DB	332	TESLDYNDYKILMAASHAFNGAHPMNGLGATSPGLGHSAPSPMGLGVMEAPLLGF	451		
QY	180	PTMNSNLSEVQKVLQIVNTVTSRQMDCKTEDISLTKYTHMKDPDSQDEEGVTSFNP	239		
DB	452	PSINSNLSEVQKVLQIVNTVTSRQMDCKTEDISLTKYTHMKDPDSQDEEGVTSFNP	511		
QY	240	VGLPVSHNGATKSIIDYTLKEVNEAKACLOSLFTDSRQISNKKELRLTLDVTDK	299		
DB	512	VGLPVSHNGATKSIIDYTLKEVNEAKACLOSLFTDSRQISNKKELRLTLDVTDK	571		
QY	300	MIENHSISTPSCQCKSPFGPIPLHQRERYLCKMNEIKAVLQPHENIVPNKAGVFD	359		
DB	572	MLESHISTPSCQCKSPFGPIPLHQRERYLCKMNEIKAVLQPHENITLTKGCVFAE	631		
QY	360	NKALLISVLSKGLTSPINPKDMSTLKAVYAMNNEPNSDELTKLSIAVGLQDFEYKE	419		
DB	632	KQALLISVLSKGLTSPINPKDMSTLKAVYAMNNEPNSDELTKLSIAVGLQDFEYKE	691		

QY	420	WFEORAKYQVXSNSSPSLERTS--KPLAPSNPTTKOSLLPRSVKMDSTSPSIAELH	477
Db	692	WFEORAKYQVXANSSPSLERTSAAEMALATLIANTTKOS--ARSPKSVDPITSSISIELH	749
QY	478	NSVTSQDPLRLIKYSHFTNIKAV-DKLDHSRSTPPLNLSTSSXNSSHSSSTYTPSFS	536
Db	750	NRVSNCTPTPLRLTKSNHPSAWKPVLDKLDHSRSTPPLNLSTSSXNSSHSSSTYTPSFS	809
QY	537	SEELQAEPLDLSTLKORREPKGIATKNTKATYSINLDHNSVSSSENSDEPLNLTYIKK	596
Db	810	SEELQAEPLDLTVLKLNESKTIATKXKSXPNIIYDHNSSVLSSETVDEPLNLTYIKK	869
QY	597	EFSNSNNLDKSNPNVFGMNFSAKPLYTPLPQSAFPAPATMPDVOTSLPGLRYPGLD	656
Db	870	EFCNANM--DKSTSPPLFGLNPFSGKPLYSAALPQSAFPAPATMPDVOTGIPGLRSTYGLD	927
QY	657	QMSFLPHAAVYTPPGAATFADMOORRYKORQOGQGLDLSAODVYMSGLDMDTSDCLS	716
Db	928	QMSFLPHAAVYTPPGAATFADMOORRYKORQOGQGLDLSOTQDYMGLDMDTSDCLS	987
QY	717	RKKIKKTESGMVACDLCDKTFQKSSLLRNKYEHTGRKPHQCQCKKAFKIKHLLIHSR	776
Db	988	RKKIKKTESGMVACDLCDKTFQKSSLLRNKYEHTGRKPHQCQCKKAFKIKHLLIHSR	1047
QY	777	LHSEKRYQCDCKGRKPSHSGSYSQHMNRKSYCKREAREEAERERAREKXGHLGPTELL	836
Db	1048	LHSEKRPQCDCKGRKPSHSGSYSQHMNRKSYCKREAREEAERERAREKXGHLGPTELL	1107
QY	837	MNRAYLOSITPQGYSDSEERESEMPRDGESEXEHEKEGEGYKGLRRRDGDEEBEES	896
Db	1108	MNRAYLOSITPQGYSDSEERESEMPRDGRRLIHEHEGGDDVYDKLRLQYGDDEEFEEBES	1167
QY	897	ENKSMODPPEITIRDEEETGDHSMDDSDGCKMETKSHDEEEMDDGM	943
Db	1168	ENKSMODPDPITIRDEEENGDSMDSDSDGCKMETKSHDEEEMDDGM	1214

RESULT 2
I50222
deltaEF1 - chicken
C/Species: Gallus gallus (chicken)
C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C/Accession: I50222
R/Funahashi, J.; Sekido, R.; Murai, K.; Kamachi, Y.; Kondoh, H.
Development_119, 433-446, 1993
A/Title: Delta-crystallin enhancer binding protein delta EF1 is a zinc finger-finger-homeodomain
A/Reference number: I50222; MUID:94116444; PMID:7904558
A/Accession: I50222
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1114 (FUN>
A/Cross-references: GB/D14313; NID:G391631; PIDN:BA03259.1; PID:G391632

[illegible][illegible]

```

RESULT 3
JX0293
zinc finger protein AREB6 - human
C/Species: Homo sapiens (man)
C/Date: 03-Feb-1994 #sequence_revision 20-Aug-1994 #text_change 28-May-1999
C/Accession: JX0293
J. Matanabe, Y.; Kawakami, K.; Hirayama, Y.; Negano, K.
J. Biochem. 114, 849-855, 1993
A/Title: Transcription factors positively and negatively regulating the Na,K-ATPase alpha
A/Reference number: JX0293; MUID:94186507; PMID:8138542
A/Accession: JX0293
A/Molecule type: mRNA
A/Residues: 1-1124 <MAT>
A/Cross-references: GB:015050; NID:9457560; PIDN:BAA03646.1; PID:01004160; PID:9558522
C/Keywords: zinc finger
F:172-193/Region: zinc finger
F:202-222/Region: zinc finger
F:242-262/Region: zinc finger
F:270-292/Region: zinc finger
F:906-926/Region: zinc finger
F:934-954/Region: zinc finger
F:962-981/Region: zinc finger
F:989-1123/Region: glutamic acid-rich

```

Query Match 33.2%; Score 1657.5; DB 2; Length 1124;
 Best Local Similarity 41.9%; Pred. No. 6,8e-79;
 Matches 406; Conservative 147; Mismatches 308; Indels 107; Gaps 32;

```

2 LTQGNKRFKTECGKAFKTKHLEHRIHSGEPYCPCKRPFSGSHSSK 61
21 VTQSCNKKFKTECGKAFKTKHLEHRIHSGEPYCPCKRPFSGSHSSK 290
62 KICGLISVNGRMNNIKTG--SSPNSVSSPTNSAITQLRNKLNGKPLMSBQTLKI 119
291 KICSLIPVNGRPRIGTKTSQCSSP-SLSASPGSPRPQRIKRIEN-KPL--QEOISVNOI 346
120 KTEPLDNDYKVLMAATHFGSSGSPFNNGLGATSPGLVHPSAQSPOHLL--GVGMEAPL 176
347 KTEPDYD--EFKPIVVASGINCSTPLONGVFTGGGGLQATSSPQGVAVVLPVGLVSPBI 405
177 LGFTMNSLSEVQKLOI-VDNTYSRQMDCKTEBISLKGYNHMKDPCSOBEGCVTSP 235
406 -----SINLSDIQVNLKAVDGNVIRQYLE--NNQANLASKQETINASPIQGGSHSV 456
236 NIPVGLPVVSHNGATKSIIDYTLKEVNEAKACLOSLTDSRRQISNIKEKRLTLDIV 295
457 -ISALSPLVODDGTITIIINYSLEQPSQLQVVPQNLKKNPVATNSCKSEKLPEDLYVK 515
236 TD-DKMEINH-SISTPFCQCFKESFPGP-PLHQHERYLCRMNEIKAVILOPHENIVPN 352
516 SEKDSFEGGVNDSTCLCDDC---PGDINALPELKH-----DLKQTPQ-----PP 560
353 KAGVVDNKKALLSVLSSEKGLTSPINPYKDHMSVLYKAYYAMNPNDELKISIANVL 412
561 LPAAEAKEPSSVSATGDNLSPOPLKNLISLKAAYALNAPSAEELSKIDSVNL 620
413 POEFVKEPFEORKVYQVNSRSPSLERTSKPLAPNSNPTTKOSLLPRSPVKPMDSTSPS 472
621 PLDVKYKFEKMQAQOIS-----VQSEBSSPEPGKV-----NIPAKNNDQPSAN 666
473 IAEIHNVSVCDDPRLRLTKSSHFTNIKAVDKLDRSNTPSPPLNLSSTSSKNSHSSSYTP 532
667 ANEPQDSTVNLQSPKMTNS---PVLPGVSTNGSSSTPSPPLNLSSTSSNTQGYLYTA 723
533 NSFSEELQAEPLDLSLPKQREPKCIATKNTKATKATSLNLDHNSVSSSENS-----DE 587
724 EG-AQEEPOVPEPLDLSLPKQGE-----LERSSTTVYQNSVSVQBE 766
588 PLNLPIKKEPFSNNLNDKSNPNVGMNPFSAKPLTYPLPQSAFPPTAMPVQTSIP 647
767 PLNLSCAKKEPOKDCVTD--SEPVVNVLPSPANPINITAIPTVTAQLPTTVAIDQNSVP 824
648 GLRPYRGLDQMSFLPHMAYTYPTGATFADMOORRYOKROGFQGLDGAODYMSGLDD 707
825 CLRALAANKQITLIPQVATYTTVSP-AVQEPPLKVIQPNQNOBERODTSSBGSVNVED 883
708 MTDSDSCLSRKIKITKTESGMVACDLCDTPOKSSSLRHKYHTGKRPQOCICCKAFKH 767
884 QMDSSTPCKKTKRKTENGMVACDLCDKIFOKSSSLRHKYHTGKRPHECICCKAFKH 943
768 KHLLEHSHLSGKRYQCDKCGKRFSGSHGSOHNNHYSYCKPAAEREAAREARX 827
944 KHLLEHSHLSGKRYQCDKCGKRFSGSHGSOHNNHYSYCKPAAEREAAREARX 999
828 GHIGPTELLMNRAYLOSTIP-OQYSDSERESMPR--DGESEKEHEKEE-----GYK 879
1000 -BAGP-EILSNHVGARASPSQG--DSDERESLTREDEDSKEKEHEKEHEKEHEKE 1055
880 LRRDQDEEEBEEBEEBESKMDTDEPTIRDEEETGDHSDSSSED-----GKET 930
1056 CERPOQDEEBEEBEEBEEBEEVE-EAENEGBEAKTEGLMKDPRASEQASLSQKVGESSE 1114
931 KSDHEEDN 938
1115 QVSEKTN 1122
    
```

RESULT 4
 A56242
 E-box-binding repressor ZEB - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 07-May-1999
 C:Accession: A56242
 M:Cell: B101, 6153-6163, 1994
 A>Title: Displacement of an E-box-binding repressor by basic helix-loop-helix proteins;
 A:Reference number: A56242; MUID:94344126; PMID:8065348
 A:Accession: A56242
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1154 <GEN>
 A:Cross-references: GB:U19969
 C:Keywords: DNA binding; zinc finger

Query Match 33.2%; Score 1605.5; DB 2; Length 1154;
 Best Local Similarity 41.1%; Pred. No. 3.7e-76;
 Matches 399; Conservative 142; Mismatches 314; Indels 115; Gaps 32;

```

2 LTQGNKRFKTECGKAFKTKHLEHRIHSGEPYCPCKRPFSGSHSSK 61
25 VTQSCNKKFKTECGKAFKTKHLEHRIHSGEPYCPCKRPFSGSHSSK 324
62 KICGLISVNGRMNNIKTG--SSPNSVSSPTNSAITQLRNKLNGKPLMSBQTLKI 119
325 KICSLIPVNGRPRIGTKTSQCSSP-SLSASPGSPRPQRIKRIEN-KPL--QEOISVNOI 380
120 KTEPLDNDYKVLMAATHFGSSGSPFNNGLGATSPGLVHPSAQSPOHLL--GVGMEAPL 176
381 KTEPDYD--EFKPIVVASGINCSTPLONGVFTGGGGLQATSSPQGVAVVLPVGLVSPBI 439
177 LGFTMNSLSEVQKLOI-VDNTYSRQMDCKTEBISLKGYNHMKDPCSOBEGCVTSP 235
440 -----SINLSDIQVNLKAVDGNVIRQYLE--NNQANLASKQETINASPIQGGSHSV 490
236 NIPVGLPVVSHNGATKSIIDYTLKEVNEAKACLOSLTDSRRQISNIKEKRLTLDIV 295
491 -ISALSPLVODDGTITIIINYSLEQPSQLQVVPQNLKKNPVATNSCKSEKLPEDLYVK 549
236 TD-DKMEINH-SISTPFCQCFKESFPGP-PLHQHERYLCRMNEIKAVILOPHENIVPN 352
550 SEKDSFEGGVNDSTCLCDDC---PGDINALPELKH-----DLKQTPQ-----PP 594
353 KAGVVDNKKALLSVLSSEKGLTSPINPYKDHMSVLYKAYYAMNPNDELKISIANVL 412
595 LPAAEAKEPSSVSATGDNLSPOPLKNLISLKAAYALNAPSAEELSKIDSVNL 654
413 POEFVKEPFEORKVYQVNSRSPSLERTSKPLAPNSNPTTKOSLLPRSPVKPMDSTSPS 472
655 PLDVKYKFEKMQAQOIS-----VQSEBSSPEPGKV-----NIPAKNNDQPSAN 700
473 IAEIHNVSVCDDPRLRLTKSSHFTNIKAVDKLDRSNTPSPPLNLSSTSSKNSHSSSYTP 532
701 ANEPQDSTVNLQSPKMTNS---PVLPGVSTNGSSSTPSPPLNLSSTSSNTQGYLYTA 757
533 NSFSEELQAEPLDLSLPKQRE-----PKGIATKNTKATKATSLNLDHNSVSSSENS 585
758 EG-AQEEPOVPEPLDLSLPKQGEELERILPPEQCLFCBGR----- 797
586 DEPLNLTKKEPFSNNLNDKSNPNVGMNPFSAKPLTYPLPQSAFPPTAMPVQTSIP 645
798 ---LEVLAKKEPOKDCVTD--SEPVVNVLPSPANPINITAIPTVTAQLPTTVAIDQNSVP 852
646 IGLRPYRGLDQMSFLPHMAYTYPTGATFADMOORRYOKROGFQGLDGAODYMSGL 705
853 VPCRALAANKQITLIPQVATYTTVSP-AVQEPPLKVIQPNQNOBERODTSSBGSVNV 911
706 DMTDSDSCLSRKIKITKTESGMVACDLCDTPOKSSSLRHKYHTGKRPQOCICCKAF 765
912 BDQNDSDSTPCKKTKRKTENGMVACDLCDKIFOKSSSLRHKYHTGKRPHECICCKAF 971
    
```

```

QY 766 KKHKHLIHSRLHSGEKPYQCDCKGRFSGSYSGOMNHRYSYCKREAEREAREAR 825
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 972 KHKHHLIHSRLHSGEKPYQCDCKGRFSGSYSGOMNHRYSYCKREAEREAREAR 1029
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 826 EKXHLGPTLLMNRAYIQSITP-QGYSDSEERESMPR--DGESEKHEKEGEE-----GY 877
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1030 ---EAGP-ELTSMENHVGARASPSQG--DSDERESLTREDEDSKKEEEDKEMBELQEE 1083
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 878 GKLRRLRDDEEESESESESESESESESESESESESESESESESESESESESESE 928
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1084 KEKEKPGQDEEESESESESESESESESESESESESESESESESESESESESE 1142
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 929 ETKSDHEEDN 938
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1143 SEQVSEKTN 1152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 5

A56037
DNA-binding protein B2P - golden hamster
C/Species: Mesocricetus auratus (golden hamster)
C/Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 05-Nov-1999
C/Accession: A56037
R/Franklin, A.J.; Jecton, T.L.; Shelton, K.D.; Magnuson, M.A.
Mol. Cell. Biol. 14, 6773-6788, 1994
A/Title: B2P, a novel serum-responsive zinc finger protein that inhibits gene transcription
A/Reference number: A56037; MUID:95021206; PMID:7935395
A/Accession: A56037
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1043 <FRA>
A/Cross-references: GB:L13856; NID:9437340; PIDN:AA37667.1; PID:9437341

Query Match 32.0%; Score 1599; DB 2; Length 1043;
Best Local Similarity 40.2%; Pred. No. 7e-76;
Matches 390; Conservative 142; Mismatches 308; Indels 130; Gaps 30;

```

QY 2 LTQAGNRKFKCTECGAKFYKHHLEHRLHSGEKPYECPCNCKRFSHSGSYSHISK 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155 VTQSGNRKFKCTECGAKFYKHHLEHRLHSGEKPYECPCNCKRFSHSGSYSHISK 214
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 62 KCTGLISVNGRNRNNTITGSSPN-SVSSPPTNSAITOLRNKLENGKRLSSEQTGLIKT 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 215 KCTSLMPVNGRPRSGLTQSCPPSPSLASPSPTRPQIRQKIEN-KPL--QEPUSVNOIK 271
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 TEPLDFNDYKVLMTATHGFSGSSPPMNG---GLG---ATSPGVNPSAGSPMQLGVGME 173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 272 TEPLDY-EFKPIVVASGINCSTPLQNGVFSGGQLQATSSPQGVVAVLPT---VGLV 326
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 174 APLLGPTNNSNLSEVQKVLQI-VDNTVSRQMDCKTEDISKLGKGMKPCSQPEQGV 232
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 327 SPI-----SINSDIQNLKVALDGNVIRQVLENNQASL-----ASKGEAV 368
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 233 TSPN-----TPPVGLPVVSHNGATKSIIDYILEVNEAKACQLGTTDSRQISNIK 284
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 369 NASLIQGGHVSISALPLVDODGTIKIINYSLEQPSQLQVVPQNLKLENPAITNSCK 428
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 285 KEKRLTLIDLVTDKMLENHSISIRPSCQFCKESFPPIPLHGERVLCKMNERIKAVLQ 344
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 429 SEKLPE--DLTVSEKKKGRPGAVDSTCLLYDCPDL-----NALPELKHYD 476
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 345 PHENIVENKAGVFNKALLLSVLSSEKGLTSPIN-PYKDHMSVTKAYAMNEPNSDEL 403
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 477 PEHPAQPPPARPEAKPEASASS--ARDGLSPSQPLKUNLISLTKAYVALNQPNSEEL 534
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 404 LKLSIANGLPOEFYKEMFEQAKTYQYSNR---SPLEKTSKPLADNSPPTKYDILPR 459
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 535 SKIADSVNLPDLVYKWKFEKQACQIPGQSEPEPSPETGTVINIPAKSDEQPOPADGSEPO 594
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 460 SPVAKPMQISPSIAELHNSVTSCDPLRLTKSGHFNINIKAVDLHDSRSNT--PSPLN 517
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 595 E-----DSAGQS-----PLKMTSS--PVLPGVSAINGRSCTSPSPSLN 633
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 518 SSTSKNSHSSSYTPNPSSEELQAEPLDLSLPKQMRPEKGIATKNTKATSIINDHS 577
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 634 SSARSIQY---FVADGAQOEPPQVEPLDSLIPKQGE-----LLERST 674
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 578 VSSSSNS-----DRLNLTFTIKKEFNSNNNDKNSNPNVFGMNPFAKPLTYPLPQSA 632
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 675 ISSVYQNSVYQVEEPLNLSVKKERQEDSCVTD--SEPVVAVIPPSANPINALPTVYA 732
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 633 PEPATMPVQTSIGLAPRYPGILDQMSFLPHMAYTYPTGAATFADMOORRYQKQFQSG 692
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 733 QLEPTVALDQNSVLCRLAANKQITILPQAYATYATVSP--ANQEPVVKYIQNGNGD 791
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 693 DLLDAQDYMSGLDMDTSDSCLSRKIKITQSGMYACDLCKTQKSSLLRHYTEHTG 752
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 792 ERQDTSSEGVSTVEDQNSDSTPPPKKTRKTENGVACDLCKIRQKSSLLRHYTEHTG 851
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 753 KRPHOCICKKAPKHHLEHRLHSGEKPYQCDCKGRFSGSYSGOMNHRYSYCKR 812
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 852 KRPHECGICKKAPKHHLEHRLHSGEKPYQCDCKGRFSGSYSGOMNHRYSYCKR 911
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 813 EAEEREAEREAREKGLHPTLLMNRAYIQSITPQGYSDSEERESMPDGESEKHEKE 872
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 912 EAEERDGTQEQEPAQ---LEALNHNHVGARASP--SQADSDERESLTREDEDSKEKE 966
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 873 GEEGYGKLRRDQDEEESESESESESESESESESESESESESESESESESESESE 930
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 967 EEE-----DKEMBELQEEKECGNAQLEEEBEE--EEHEEMDAKDEAAKTGAIVEN 1015
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 931 KSDHEEDNME 940
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1016 GAAPQAGSLE 1025
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 6

JC4934
delta-erythralin/E2-box factor 1 - mouse
C/Species: Mus musculus (house mouse)
C/Date: 22-Oct-1996 #sequence_revision 01-Nov-1996 #text_change 05-Nov-1999
C/Accession: JC4934
R/Sekido, R.; Takagi, T.; Okanami, M.; Moribe, H.; Yamamura, M.; Higashi, Y.; Kondoh, H.
Gene 173, 229-232, 1996
A/Title: Organization of the gene encoding transcriptional repressor deltaE1 and cross-
A/Reference number: JC4934; MUID:97082972; PMID:8964504
A/Accession: JC4934
A/Status: preliminary; nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1-1117 <SRK>
A/Cross-references: DDBJ:D76432; NID:91027499; PIDN:BA11177.1; PID:dt011838; PID:910275
A/Comment: This protein is a transcriptional repressor which binds at the E2-box sequence
omain.
C/Genetics:
A/Genes: deltaE1
A/Keywords: zinc finger
F/150-277/Region: zinc fingers
F/882-964/Region: zinc fingers

Query Match 31.8%; Score 1587.5; DB 2; Length 1117;
Best Local Similarity 39.9%; Pred. No. 3.1e-75;
Matches 395; Conservative 142; Mismatches 307; Indels 145; Gaps 32;

```

QY 2 LTQAGNRKFKCTECGAKFYKHHLEHRLHSGEKPYECPCNCKRFSHSGSYSHISK 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 211 VTQSGNRKFKCTECGAKFYKHHLEHRLHSGEKPYECPCNCKRFSHSGSYSHISK 270
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 62 KCTGLISVNGRNRNNTITG--SSPNVSSPPTNSAITOLRNKLENGKRLSMESEQTGLIKT 119
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 271 KCTSLMPVNGRPRSGLTQSCSP-SLSTSPSPPTRPQIRQKIEN-KPL--QEPUSVNOI 326
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 120 TEPLDFNDYKVLMTATHGFSGSSPPMNGGLGATSPLGVPASQSPKQHL---GVGMEADL 176
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 327 KTEPLDY-EFKPIVVASGINCSTPLQNGVFSGGQLQATSSPQGVVAVLPTVGLVSP 385
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 177 LGPPTNNSNLSEVQKVLQI-VDNTVSRQMDCKTEDISKLGKGMKPCSQPEQGVTS 235
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 386 -----SINLSDIQVLYKVAADVGNVIRQVLETNQ---ASLASKQEQAVASASPIQQGCHSV 436
Qy 226 NIPVGLPVVSHNGATKSIIDYTLLEKVEAKACLSLTDSRQISNIKEKLRLLIDLV 295
Db 437 -ISAISLPLVODDGTTKIILINSLQPSOLOVVPQNLKEIPAPINSCSEKLPEDLVK 495
Qy 296 TD-DKMIEN-HSISTPFGQCFKESFPG---PIPLHQRVYLCKMNEIKAVLQPHENIV 350
Db 496 SETDSFGARDSTCLCEDC---PGDLNALPELKHVDPECPAQPPAPATKSPSS 551
Qy 351 PNKAGVFPDNKALLISVLSSEKGLTSPIN-PYKDMNVULKAYYANMPENSDELLKISIA 409
Db 552 ASSAG-----NGDLSPSQPLKNLILSLKAYYALNAQPSLTELKIDAS 595
Qy 410 VGLPOEFYKEMFEQRKVQYNSRSPS-----LERTSKPLAPNSPTTKDSL 457
Db 536 VMLPLDGVKAFKFKQAGQIP-GQSPDPSPGCTGVNIPTKTQOPADGNEPQEDSTR 654
Qy 458 PRSPVYPNDSTSPSIAELHNSVTSCDPLRLTKSSHFTNIKAVDKLDHSRNTSPSLNL 517
Db 655 GQSPYKIHSSPLP-VGSAMNGSRSC-----TSSPSPLNL 688
Qy 518 STSSKSNHSSSYTNSFSSEELQAEPLDLSLPKQMRPEKGIATKNTKATSNILHNS 577
Db 689 CSARPPQGS---CVAEQAQEEPOVEPLDSLPKQGE-----LLERST 729
Qy 578 VSSSENS-----DEPLNLTFIKKEFSNNLNNKSNPVFGMNPFAKELTPLPQSA 632
Db 730 VSSYQNSVYVQSEPLNLSCKKEPQKDSVTD--SEPVVAVVPANPINALPTVA 787
Qy 633 PPATFMPVQTSIIGLRPPYGLDQMSFLPHAYVYPTGAATFADMOQR--KYORKQGF 690
Db 788 QLPITVALADQNSVPCALANKQITILLQVAYTY---SATVSAVQEPYKVIQPNGN 844
Qy 691 QGLDLGAQDYVSGLDMDTSDSCLSRKKIKTSGEMTACDLCDTPOKSSLLRHKTEH 750
Db 845 QDERDSTSEGVSTVEDQNDSDSTPPKKTREKTENGMAACDLCDFQKSSLLRHKTEH 904
Qy 751 TGRPHQOICCKAKKHKHLIHSRLSGEKPYQDCGKRFSGSYSQMNNRYSYC 810
Db 905 TGRHHEGICCKAKKHKHLIHSRLSGEKPYQDCGKRFSGSYSQMNNRYSYC 964
Qy 811 KREAEREAEREAREKKGHPTEL--LMNRAVYLSITFGYSDSEERESMPR--DGSE 866
Db 965 KRGAEDRAMGE-----DAGPEVLPYVATATHVGAARASPSQADDERESLREDEDESE 1019
Qy 867 KEHEKEGEE---GYGK-LRRRDGDEEEEEESEENKSMQTDPTTIDDEETGDH--- 917
Db 1020 KEEBEDKEMELQCGKCEENPQGESEEEEEESEEEVEADEAHEBAAKTDGT 1079
Qy 918 -----SMDSDSEDKMETKSDHEE 936
Db 1080 VEVGAQAQAGSLQKASSESESESESEQ 1108

```

RESULT 7

S33641
 homeotic protein zfh-1 - fruit fly (*Drosophila melanogaster*)
 C:Species: *Drosophila melanogaster*
 C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 24-Sep-1999
 R:Accession: S33641; S27816
 R:Portint, M.E.; Lai, Z.; Rubin, G.M.
 Mech. Dev. 34, 113-122, 1991
 A:Title: The *Drosophila* zfh-1 and zfh-2 genes encode novel proteins containing both zinc
 A:Reference number: S33641; MUID:92001539; PMID:1680376
 A:Accession: S33641
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1060 <FOR>
 A:Cross-references: EMBL:M63449; NID:g158820; PIDD:AAA23050.1; PID:g158821
 C:Genetics:
 A:Gene: zfh-1

A:Cross-references: FlyBase:FBgn0004606
 C:Superfamily: unassigned homeobox proteins; homeobox homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation; zinc finger
 F:706-762/Domain: homeobox homology <Hox>

Query Match 16.4%; Score 819.5; DB 2; Length 1060;
 Best Local Similarity 29.2%; Pred. No. 4e-35;
 Matches 256; Conservative 100; Mismatches 268; Indels 253; Gaps 33;

```

Qy 9 RRFKTECGKAPKYVGHILKEHLRHSGEKPYECPNCKKRFSGSYSSHSKCGICGIS 68
Db 359 RRFKTECGKAPKYVGHILKEHLRHSGEKPYECPNCKKRFSGSYSSHSKCGICGIS 415
Qy 69 VNGMRKNN-----IKTSSPNVSSSPTNSAITOLANKLEN-----GKPLSSEQTGLKTI 119
Db 416 MGLKLNRRALLKRLKESPGSASASRRSPDHGKGLPEQPSLGLPHNS----- 467
Qy 120 KTEPLDFNDYKVLATHGFGSSGSPMNGLGATSL-GVHPSAGSPMCHLVGMEAPLIG 178
Db 468 -----YFASDAQVQGSAAAPAPFPFHPY-----NMAALIA 499
Qy 179 FP-----TMSNLS-----EVQVLTQIVDNTVSRQKMDCKTEDISKLGYNKDCSQPE 228
Db 500 FPHNFMALAGLDPRVHPYSIORLLQ--SAAGQQREREERQKQOQ--DEETPD 554
Qy 229 EGVTSFNI PPVGLPVVSHNGATKSIIDYTLLEKVEAKACLSLTDSRQISNIKEK 288
Db 555 EPKLVMDIEEBPTKEMATPPTATEAATPIKRESEEA-----SPREYTRSSQAIKQEO 609
Qy 289 RTLLIDLVTDKMIENH--ISTPFGQCFKESFPPIPLHQRVYLCKMNEIKAVLQ 345
Db 610 EPLVAERQRTVEEHAADVHESADLRCSGKQFNHPELVQHEKVLGL--IKEELQ 666
Qy 346 HENIVPNKAGV-----DNKALLSVLSEK-----LSPINPYKDMNSVLKA 390
Db 667 H-FOOQATSPALASAEDEDEDEMDVEEPRQESSEKRYVTAI--EEQOOLQ 722
Qy 391 YYANMPENSDELKISIAVGLPOEFYKEMFE-----ORKVQYNSRS----- 434
Db 723 HYSLNARSRBEFRILARQLDPRVQVWFQNNRERKQSPQNNQAAGAAMPID 782
Qy 435 -PSLERTSKPLAPNSPTTKDSLPRSPVVKPMDSTSPSIAELHNSVTSCDPLRLTKSS 493
Db 783 QASLTRDQPL-----DLVGRQDPLTPKSESPYIA-----PP-----SG 818
Qy 494 HFTNIKAYD-KLDHSRNTSPPLNLSSTSSKSHSSTTPNSFSSEELQAE----- 544
Db 819 EALNPEALNLSRKFTSASMSPASISPSAALYGAAPRPSPPNSQLDSTPRSGAEPG 878
Qy 545 -----LDSLPRQ-----REEPGIIATKNTKATSIINL-DHNSVSSSENSDEPLNLTPI 594
Db 879 LPVYVLPMSLPEALFKR-PEGDFAS-NHALMSIKLPDTRGTSLSFGSGP----- 928
Qy 595 KKEFSNNLNNKSNPVFGMNPFAKELYPLPQSAFPATFMPVQTSIIGLRPYPG 654
Db 929 KRSWRDD-----SRISHEDEFGAGVLMRP----- 953
Qy 655 LDQMSFLPHMAYVPTGATPADMQRRKYORKQFQGLDLDAQDYVSGLDMDTSDSC 714
Db 954 -----KPRRGKVEYTHHAD-----PDLP----- 972
Qy 715 LSRKKIKTSGOMVACDLCDTPOKSSLLRHKTEGKRPHQOICCKAFKGHKHLI 774
Db 973 -----YCDQCDKAFKQSSLLAKKIKHSQRPQCIETCFKAFKGHKHLI 1019
Qy 775 SRLHSGEKPYQDCGKRFSGSYSQMNNRYSYCK 811
Db 1020 KRLHSGEKPYQDCGKRFSGSYSQMNNRYSYCK 1056

```

RESULT 8

T29204
 hypothetical protein F28F9.1 - *Caenorhabditis elegans*

C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T29204
 R:Neelson, J.; Wohlmann, P.
 submitted to the EMBL Data Library, September 1996
 A:Description: The sequence of C. elegans cosmid f28f9.
 A:Reference number: Z20587
 A:Accession: T29204
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-680 <NEU>
 A:Cross-references: EMBL:U70850; PIDN:AA09122.1; GSPDB:GN00022; CESP:F28F9.1
 A:Experimental source: Strain Bristol N2; Clone f28f9
 C:Genetics:
 A:Gene: CESP:F28F9.1
 A:Map position: 4
 A:Introns: 53/3; 234/3; 268/3; 331/3; 485/2; 531/3; 589/1

Query Match 14.8%; Score 741; DB 2; Length 680;
 Best Local Similarity 25.6%; Pred. No. 2, 7e-31;
 Matches 228; Conservative 106; Mismatches 217; Indels 338; Gaps 27;

```

OY 9 RRFKTECGKAPKFKYKHLKHELRHSGEKYRPECNCKRPSHSGSYSSHSKCCIGLIS 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 22 RRFKCECTKAFKFKHLKHELRHSGEKYRPECQCKRPSHSGSYSSHSKCCV---- 77
OY 69 VNGRMNNIKTGSSPNVSSPTNSAITOLRNKLENGKPLMSBQGTGLKIKTEPLDFND 128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 78 -----QQASPSMVT-----PFPD 90
OY 129 YKVLATHTGFGSSGPPMNGGLATSPGLVHPAQSOMHLCVGEAPLIGF-PTMNSNIS 187
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 91 YQLMM-----YRNIMLOLQFPQVSLFSTANNN 119
OY 188 EVQKVLQIVDNTVSRQMKDCTEDISKLGYNHKDCSQPEEGVTSPIPIGLVYSH 247
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 120 DVMSLQ-----ANLFQSLNGTSPTTQEPSADSPKLEVVDEEVS- 164
OY 248 NGATKSIIDYLEKVNKAQLQSLTTSRQISNIKEKLRILDLVTDKKIENHSIS 307
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 165 -----SEVKTE-VKT--EVKTEDSVPE-ESIT 187
OY 308 TFPSCQCKESFPGPPIPLQHERRYLCQMEELKAVLQPHENIVPNKAGVFNKA----- 362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 188 PAVSMEL-----SPAP-----EQNGNESMNGGSGDGKSSPDWR 222
OY 363 -LLLSVSEKGLTSPINPKDMSTLKAYANMMPNSDELKISIAVGLPQEPFKEMF 421
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 223 PLRSRFLND-----SQVAVLQNHFKKNPFPSKTELSAVAQIGVNRVVQVWF 271
OY 422 EQRKVVQYNSRSPLERTSKPLAPNSNPTTKDSLPRSPVKPMDSITSPSIAELANSVT 481
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 272 QNTRAKERSNRLPSNPRGSVASAAA-----AATSPTWQ----- 307
OY 482 SCDDPLRLTK--SSHFTNIKAVDKLDHSRNTSPPLNLSSTSSKNSHSSSYTPNSFSEE 539
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 308 -----TPVQLMAAAMASQFSN-----GNNSLTASQDRNNMENTDEAVDHQ 347
OY 540 L---QAEPLDLSPKQMRPK-----GIATKTKTKAIS 570
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 348 LKDGKPTPLDLTSTDTDEPMSPEKLIGLDQYGVIOELLRQAGNGFATQDEDEEKP 407
OY 571 INLDHNSVSSSEN-----SDEPLNLTFIKKEFSNNSNLDKSN 610
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 408 IKAEEPSVSSGSSSIMPSTIGQYPSILDSASLSVLEKALDVST--DFS----- 454
OY 611 PVFGMNPFAKPLPYTLPPOSAPPPATFMPPVQTSIPGLRYPGLDOMSLFPMAYTPT 670
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 455 -----ECFAIQFLMLHVLNVSVMK-----EQKTL-----AYLS 484
OY 671 GAATPADMOQRK-----YQKQKQFG-----DLIDGAQYVMSGLDMMTSDSC 714
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 485 GKNFDSIQFREDIRAVRVYIYSNKQIFPKLYPPRYALTNNFODMOCKSSSDASILC 544

```

RESULT 9
 S35305
 zinc finger protein ZNF91 - human
 C:Species: Homo sapiens (man)
 C:Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 05-Nov-1999
 C:Accession: S35305
 R:Bellefroid, E.J.; Marine, J.C.; Ried, T.; Lecocq, P.J.; Riviere, M.; Anemiy, C.; Ponce
 EMBL J. 12, 1363-1374, 1993
 A:Title: Clustered organization of homologous KRAB zinc-finger genes with enhanced expres
 A:Reference number: S35305; MUID:93223677; PMID:8467795
 A:Accession: S35305
 A:Molecule type: mRNA
 A:Residues: 1-1191 <BEU>
 A:Cross-references: EMBL:L11672; NID:g186773; PIDN:AA59469.1; PID:g186774
 A:Note: the authors translated the codon GCA for residue 750 as Thr and GCT for residue 1
 C:Genetics:
 A:Gene: GDB:ZNF91; HPF7; HTP10
 A:Cross-references: GDB:132284
 A:Map position: 19p12-19p12
 C:Keywords: DNA binding; zinc finger

Query Match 7.6%; Score 379; DB 2; Length 1191;
 Best Local Similarity 21.5%; Pred. No. 5e-12;
 Matches 214; Conservative 115; Mismatches 365; Indels 300; Gaps 39;

```

OY 7 GNRKFKTCGKAFKFKYKHLKHELRHSGEKYRPECNCKRPSHSGSYSSHSKCCIGL 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 290 GEKPYCECGKAFSHSSTLAKHRIHTEKPYKCECGKAFSHSALAKH----- 340
OY 67 ISVNGMRNNIKTGSSPN-----SVSSPTNSAITQLNNK----- 101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 341 -----KRHTGEKPYKCECGKAFSHSSTLANHKTITHEKPYKCECGKTFGRSLT 392
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 102 -----LENGKPLMSBQGTGLKIKTEPLDFNDYKVLATHTG--SGSSPPMNGGLA-- 151
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 393 LTKHKITIHAGKLYKCECGKA-----FRRSSNL-TIHKFIHTEKPYKCECGKAF 443
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 152 --TSPGVNHP--SAQSPMQLGCVGEAPLIGFPTMNSNLSEVQKVLQIVDNTVSRQMD 206
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 444 NMSSSLTKHHRFHTREKPFKCECGK-----GF-----TMSSTLTTHKRI 483
OY 207 CTEDISKLGYNHMDPCSQPEEGVTSPIPIVGLPVVSHNGATKSIIDYLEKNEAK 266
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 484 HTGE-----KPYKCE--CGKAFRQST-----LTKHKITIHTEKPYKE 521
OY 267 ACIQSLTTSRQISNIKEKLRILDLVTDKKIENHSISTPFCQCKESFPGPPIPLA 326
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 522 ECGKAF-----RQSLTLNKGKI-----HSREKPYKCECGKAFKQFSTLT 562
OY 327 QHE-----RYLCQMEELIAY-----LQPHENIVPNKAGVPD--NKALLSSVL-SEK 373
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 563 THKTIHAGKLYKCECGKAFNHSSTLTHKTIHTEKPYKCECGKAFKQFSTLT 622
OY 374 LNSPINPK-----DMSVLAAYANMMPNSDELKISIAVGLPQEPFKEMFQEK 425
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 623 IHTGEKPYKCECGKAFSHSALAKH-----KRHTGEK 656
OY 426 VYQ-----YNSRSPLETS-----KPLAPNSPTTKDSLPRSPVKPMDSITSPSIA 474
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 657 PYKCECGKAFNSSTLANHKTITHEKPYKCECGK-----FKRLSTLTGKII 707
OY 475 ELHNSVSCDP-PLRLTKSSHFTNIKAVDKLDHSRNTSPPLNLSSTSSKNSHSSSYTPN 533
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


Db 708 HAGEELVYCEECGKAFNRSSNLT-----IHKFIHGE---KPYKCECGKAFNWSSTLTGH 760
 QY 534 S--FSSEELQAEPLDLSLPKQMR- PKGIATKNTKATSIINDHNSVSSSENDPELN 590
 Db 761 KRIHREK-----PFKCECGKAFIWSSTLTIRKRIH-----TGEEKYK 799
 QY 591 LTFIKKESNNNDKNSNNPFGMPPSAKPLVPLPQSAFPATMPPVQTSIPGLR 650
 Db 800 CEECGKAFSSSTLT-TKHTIHTGEEKPYKCEC-----GKAFKISSALAKKHIIHAGEK 852
 QY 651 PY-----PGIDQWFLPHMATYPTGAATPAD-----MOQRKTKQKQKQ 692
 Db 853 LYKCECGKAFNQSNTLTTHKIIHTKEKPSCECDKAFIWSSTLTTEHRIHTREKPYK 912
 QY 693 DILDQADQYMSGL-----DDMTDSCLSRKKIKTESGMVACDCLD 734
 Db 913 EECGKAFQPSHILTHKMHTEGEEKPYKCECGKAFSSSTLTTHKIIHTGEEKPYKCECG 972
 QY 735 KTFQKSSLLRHKYHTGKRRPHQCI CKKAFKHKHLLIHSRLHSGEEKPYQCDKGRPS 794
 Db 973 KAFKRSSTLTTEHRIHTGEEKPYKCECGKAFSSSTLTFRHTMHTGEEKPYKCECGKAFN 1032
 QY 795 HSGSISQMHNRVSYCKREAEERAEAREKHLGPELLMNPAYLOSTIPQGYSDS 854
 Db 1033 RSSKLTTHKIIHTGEEKPYKCEC-----CGKAFISSSTLTNGHKRIH 1072
 QY 855 ERESMPRDESEK-----EHEK--EGEEGY 877
 Db 1073 TREKPYKCECGKAFSSSTLTIRKRIHTEGKY 1106

RESULT 10

500647
 finger protein - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C>Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 31-Dec-1993
 C:Accession: S00647
 R:Ruiz J. 6, 3065-3070, H.; Perry-O'Keefe, H.; Melton, D.A.
 EMBL J. 6, 3065-3070, H.; Perry-O'Keefe, H.; Melton, D.A.
 A:Title: Xfin: an embryonic gene encoding a multifingered protein in Xenopus.
 A:Reference number: S00647; MUID:88082679; PMID:2826129
 A:Accession: S00647
 A:Molecule type: mRNA
 A:Residues: 1-1350 <RUI>
 A:Cross-references: EMBL.X06021
 A:Note: it is uncertain whether Met-1, Met-11, Met-16, Met-38, or Met-39 is the initiator
 C:Genetic: A:Gene: Fln
 C:Keywords: DNA binding; zinc finger

Query Match 7.2%; Score 358; DB 2; Length 1350;
 Best Local Similarity 19.3%; Pred. No. 7.4e-11;
 Matches 189; Conservative 135; Mismatches 338; Indels 318; Gaps 39;
 QY 6 AGNRKFKTECGKAFKYLKHLKHLRHSIGKEPYECPPKRRFSSGSSSHISSKCIIG 65
 Db 526 SGEKKFQCAECCKGKTOKSDLVKHIRVHTGEEKPFCLCKSSFSQNSDLHKTM----- 578
 QY 66 LISVNGRMNRNITKSSSPNSVSSSPTSNAITOLRNKLENGKPLMSBOTGLIKITETPLD 125
 Db 579 -----RIHTEKEFPCTYCK-----SPTERSALIKHRTHTG 611
 QY 126 FNDYKVLMAHTFGSSSPFMNGGLGATSPIGVHPASQPMOHLGVGMNAPLLGPTTMNSN 185
 Db 612 ERPHKCSYQCKF-----IQXSALTGHSTHTG-EKP--YFCTCG 649
 QY 186 LSEVQKVLQIVNTVSRQK-----DCK--TEDISKLKGYHMK-----PCS 225
 Db 650 KSFIGN-----SDLVKQRIHTGEEKPYCTENKRFSSGLV--HRTSGEEKPYRCP 702
 QY 226 QPEEGVTSNPVPGVPLVSHNG-----ATKSIIDITTEKVNKA-----C----- 268
 Db 703 QCEKTFIQSSDLV--HLVYVHNGENPPAATFAHEILIRRENTLRSEBDPYCTECGVF 759

QY 269 -----LOSLTDSRRQISNIKKEKRLTLDLVTDKMIENHSISTPSCQFCESF 319
 Db 760 HQRPALLKHLRHTHTKRYPCNECDKSPQI-SDLV---KILRHTGGRPHCEBCKNGF 815
 QY 320 PGIPLHOHER-----YLCRMBE-----IKAVLOPHENIVPNKA-----GVFVNDK 361
 Db 816 IONSDDLVGQHTHTGERPYTSCQCDKGIQRSALTGHRTHTGEEKPYKCECQCKFIQNS 875
 QY 362 ALLL-----SSVSEKGLTSPINPYK-----DHMSVKA 390
 Db 876 DLVHQRHTGEEKPYHCPDCKRFTGSSSLIKHQRHSRIKPYPGVCGKSFSSSNTLK 935
 QY 391 YYAMNMEPNSDELKISIAVGLPOEFVKEWQKRVQYSSNRSPSLERTSKPLAPNSNP 450
 Db 936 HLKCHSENP-----PALSEIGFVAE-----TQTHP 963
 QY 451 TTQDSL--PRSPVKPMD-----ITSPSIAELHNSVTSQDPLRLTKSSH 494
 Db 964 DPVHIVYGDYASYISPEAAGERSFKNDGKCPAHSVLIKHVRHTGERPYKSCQCTR 1023
 QY 495 FTNIKAVDKLHRSNTPSPINLSTSSKNSHSSYTPNFSSELOAEPLDLSLPKQMR 554
 Db 1024 -SFIOKSDLVKHYVHT-----GERPYKCGLCERSFVEKS-----ALSRQR 1064
 QY 555 EPKGIATKNTKATSIINDHNSVSSSENDPELNLT-----FIKKEFSNNINDKNSN 610
 Db 1065 -----VHKNSPVLNSAMEQOQTYNGESKDQNSLVPLQHLVKE--SEPHIVN----- 1112
 QY 611 PVFGMNPSPAKPLVPLP-POSAPPATMPPVQTSIPGLRPGLDQMSFLPHMATYTP 669
 Db 1113 -----AVPSLILQSYFP-----ILHP 1130
 QY 670 TGAATFADMOORRKYQKQKGF--OGDLDGADQYMSGLDMDTDSCLSRKIK-KTESG 726
 Db 1131 KGTFRYSCEGKCFTHRSVFLKIMRMHTTGQPTCKEKGKSFQSSALVGHVRIHTGEK 1190
 QY 727 MYACDLCKTFOKSSLLRHKYHTGKRRPHQCI CKKAFKHKHLLIHSRLHSGEEKPYQ 786
 Db 1191 PYACSTCKSTIQOSDLAKQRIHTGEEKPYCTYCGKKFIDRSSVVGKSRHTGERPKC 1250
 QY 787 DKCGRRFSHSGSYSOHNN-----HRSYCKREAEERAEAREKHLGPELLMNR 839
 Db 1251 NECTKGFGQKSDLVKHMRTHTGEEKPYGNGCCDRFSFTHSAVRQR-----MCN- 1299
 QY 840 AYLOSTIPQGYSDSEERBSM 859
 Db 1300 -----TGRPYQDEBYENSL 1313

RESULT 11

541705
 EVI1 protein - human
 C:Species: Homo sapiens (man)
 C>Date: 25-Dec-1994 #sequence_revision 10-Nov-1995 #text_change 17-Mar-1999
 C:Accession: S41705
 R:Mitani, K.; Ogawa, S.; Tanaka, T.; Miyoshi, H.; Kurokawa, M.; Mano, H.; Yazaki, Y.; Ohn
 EMBL J. 13, 504-510, 1994
 A:Title: Generation of the AML1 - EVI-1 fusion gene in the t(3;21) (q26;q22) causes blasti
 A:Reference number: S41704; MUID:94147997; PMID:8313895
 A:Accession: S41705
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1042 <MIT>

Query Match 7.1%; Score 356.5; DB 2; Length 1042;
 Best Local Similarity 19.5%; Pred. No. 6.2e-11;
 Matches 205; Conservative 144; Mismatches 402; Indels 301; Gaps 36;
 QY 9 RKFKTECGKAFKYLKHLKHLRHSIGKEPYECNCKKRRFSSGSSSHISSKCIIGLS 68
 Db 101 REYKDCQPKAFNWKSNLIRHQMSHSGKHGECENCAKAVFTDPNSLQRHRSQHV----- 155

```

QY 69 VNGRMNNIKTGGSPNSVSSPTNSAITQLRNKLENGKPL-----SMSEOTGLIKIKT 121
DB 156 -----GARAHACECGKTFPSSGLKQHKHHSVKRFCEVCHKSYTQPSNLCRHRR 208
QY 122 EPLD-----FNDYKYLMAATHGF-SGSSFPNNGIGLATSPLGVHPSAOSPMOH 167
DB 209 MHADCTQIKCKDCGCMFSTTSSLNKRFRFCGKNHFAAGF-----250
QY 168 LGVMEAPLLGFPPTMNSNSEVQKVLQIVDNVTSRQMKDCTEDISKLGYNHKDCPSQP 227
DB 251 FGGGISLP--GTPAMDK-----TSMVNMHANGGLADYTG-----ANR 286
QY 228 EEOGVTSPNIP-----PVGL-----PVVSHNGATSIIDYTLKYNKAEKACIQSLT 273
DB 287 HPAGLFTPTAPGFSFSPGLFPGLYHRPPLIPASSPVKGL--SSTEQTNKSSQ-----338
QY 274 TDSRRQISNIKKKELATLIDLVTDDKMIENHSISTFSCQFCSEFPGRPLPHQHRYL 333
DB 339 -----PLMTHPQILP 348
QY 334 KMMBEIKAVLQPHENIVPNKAGVFDNKALLSSVLSSEKGLTSPINPYKDHMSVLKAYYA 393
DB 349 ATODILKA--LSKHPISGDK-----PVELQPERSSSEERPFER 384
QY 394 MNMEPNSEDLKLSIAVGLPQEFVKEMFEQRKVVQYNSNSRPSLERTSKPLAENSPTTK 453
DB 385 ISQSESSDLDDVSTPSGDLFTT-----SGSDEJESDIESDKKFKENGKMF 432
QY 454 DSLPPSPVKPM-----STSPSIAP--LHNSVTSQDPLRLTSSHTNINIKAVD 502
DB 433 DKVSPLONLASINNKKEYSNHSIFSPSLEQTAVSGAVNDISIKALISAEKYGSGTLVG 492
QY 503 KLDHRSNTPSPPLNLSSTSSKNSSSYTPNSFSSELOAEPLD-----SLPKOM 553
DB 493 LQDKKYGALPYR--SMPLPFPFPAFSQSMYR--FPDRLSLPLKMEPOSBEVKYLOKS 549
QY 554 REBKGIATKNTKATKATISINDHNSVSSSENSDEPLNTFIKKEFSNSNLDNKSNNPV 613
DB 550 SESEPLUTTKRDEKELTPVPSKRPVTPATSOQDPLDSGSRASGTLTEPRKNHVF 609
QY 614 G-----MNPESAPLYTLPPLPQSAFPATMPV-----QTSIP-----GLRPY 652
DB 610 GKKKGSNVEPRPADSL-----QHARPTPFMDPIYRVEKRLTDPLEALKEXYLRS 663
QY 653 PGL-----DQMSFLPHMAVYTPGAATFADMQQRKRYQKQFGODLLDGAQDVMSGLDDMT 709
DB 664 PGLFHPQMSAIENTMA-----EKLSEFSA--LKEPASELLQSVPSMFR 703
QY 710 D-----SDSCLSRKKIKTESGMVACDLCDKTFQKSSSILRHKYBHTGRPHQCOICCKAFK 766
DB 704 NFRAPPAALPENLRLRKGKE--RYTCRYCGKIFPRSANLTRILRTHGTGQPYRCKYCDRSFS 762
QY 767 HKHHLIENSR--LHSGEPYQCDCKGKRFSGSYSQM--NHRYSYCKREAEEREA 824
DB 763 ISSNLQRAHVNIHNKEKPFKCHLCDRCFGQOTNLDRLKKGENGNSGTATSSPHSELES 822
QY 825 -----REKHLGP--TELLANRAYLOS-----ITPOGYS--DSE 854
DB 823 TGAILLDDKEDAYTEIRNFIGNSNHGSQSPRANVERNNGSHFKDEKALVTSQNSDLDDE 882
QY 855 ERESMPDGESEKEHEKEGEYQKLT--RRDGDDEEEEBEESENKSMJDTPETIRDEE 912
DB 883 EYDEDEVLLDBEDBDNDITGKTGKEPVTSNLHEGNPEDDYETSALMEWSCITSFVRYKEEE 942
QY 913 -ETGDHSM-----DSEEDGKMETKSDHEED 937
DB 943 YKSGLSALDHIRHFTDLSLKKRMEEDNQYSEAE 974

```

RESULT 12

A60191
oncogene Evi-1 - human
C:Species: Homo sapiens (man)

C>Date: 20-Feb-1993 #sequence_revision 17-Apr-1993 #text_change 05-Nov-1999
C/Accession: A60191
R/Morishita, K.; Parganas, E.; Douglass, E.C.; Ihle, J.N.
Oncogene 5, 963-971, 1990
A>Title: Unique expression of the human Evi-1 gene in an endometrial carcinoma cell line:
A/Reference number: A60191; MUID:90326419; PMID:2115646
A/Accession: A60191
A/Molecule type: mRNA
A/Residues: 1-1051 <MOR>
A/Cross-references: GB:X54989; NID:950873; PIDN:CAA38735.1; PID:950874
A/Note: the authors translated the codon CCT for residue 85 as Leu
C/Genes:
A/Name: GDB:EVI1
A/Cross-references: GDB:119889; OMIM:165215
A/Map position: 3q26-3q26
C/Keywords: alternative splicing; DNA binding; zinc finger

Query Match 7.1%; Score 355.5; DB 2; Length 1051;

Best Local Similarity 19.4%; Pred. No. 7.1e-11; Matches 199; Conservative 140; Mismatches 395; Indels 293; Gaps 36;

```

QY 9 RKRCTEGKAPFYKHLKELRHSGEPRYECPCNCKKFSHSGYSYSHSKKICIGLS 68
DB 101 REYKDCQCKAFRWKSNLIRHQMSHDSGKHYECNCACAKVFTDPSNLRHTRGQV-----155
QY 69 VNGRMNNIKTGGSPNSVSSPTNSAITQLRNKLENGKPL-----SMSEOTGLIKIKT 121
DB 156 -----GARAHACECGKTFPSSGLKQHKHHSVKRFCEVCHKSYTQPSNLCRHRR 208
QY 122 EPLD-----FNDYKYLMAATHGF-SGSSFPNNGIGLATSPLGVHPSAOSPMOH 167
DB 209 MHADCTQIKCKDCGCMFSTTSSLNKRFRFCGKNHFAAGF-----250
QY 168 LGVMEAPLLGFPPTMNSNSEVQKVLQIVDNVTSRQMKDCTEDISKLGYNHKDCPSQP 227
DB 251 FGGGISLP--GTPAMDK-----TSMVNMHANGGLADYTG-----ANR 286
QY 228 EEOGVTSPNIP-----PVGL-----PVVSHNGATSIIDYTLKYNKAEKACIQSLT 273
DB 287 HPAGLFTPTAPGFSFSPGLFPGLYHRPPLIPASSPVKGL--SSTEQTNKSSQ-----338
QY 274 TDSRRQISNIKKKELATLIDLVTDDKMIENHSISTFSCQFCSEFPGRPLPHQHRYL 333
DB 339 -----PLMTHPQILP 348
QY 334 KMMBEIKAVLQPHENIVPNKAGVFDNKALLSSVLSSEKGLTSPINPYKDHMSVLKAYYA 393
DB 349 ATODILKA--LSKHPISGDK-----PVELQPERSSSEERPFER 384
QY 394 MNMEPNSEDLKLSIAVGLPQEFVKEMFEQRKVVQYNSNSRPSLERTSKPLAENSPTTK 453
DB 385 ISQSESSDLDDVSTPSGDLFTT-----SGSDEJESDIESDKKFKENGKMF 432
QY 454 DSLPPSPVKPM-----STSPSIAP--LHNSVTSQDPLRLTSSHTNINIKAVD 502
DB 433 DKVSPLONLASINNKKEYSNHSIFSPSLEQTAVSGAVNDISIKALISAEKYGSGTLVG 492
QY 503 KLDHRSNTPSPPLNLSSTSSKNSSSYTPNSFSSELOAEPLD-----SLPKOM 553
DB 493 LQDKKYGALPYR--SMPLPFPFPAFSQSMYR--FPDRLSLPLKMEPOSBEVKYLOKS 549
QY 554 REBKGIATKNTKATKATISINDHNSVSSSENSDEPLNTFIKKEFSNSNLDNKSNNPV 613
DB 550 SESEPLUTTKRDEKELTPVPSKRPVTPATSOQDPLDSGSRASGTLTEPRKNHVF 609
QY 614 G-----MNPESAPLYTLPPLPQSAFPATMPV-----QTSIP-----GLRPY 652
DB 610 GKKKGSNVEPRPADSL-----QHARPTPFMDPIYRVEKRLTDPLEALKEXYLRS 663
QY 653 PGL-----DQMSFLPHMAVYTPGAATFADMQQRKRYQKQFGODLLDGAQDVMSGLDDMT--710
DB 664 PGL-----PLFHPQFQILPDDORTMNSAIENTMA--EKLSEFSA--LKEPASELLQSVPSMFR 715

```

QY 711 -SDCLSRKKIKKTESGMVACDLCKTFQKSSSLRHKTEYTKGRPHQCICCKAFKHKH 769
 Db 716 APPNALPMLLJKGKE-RYTCRYCGKIFPRSANLTRHLRTHGEOPYCKYCDRFSFSS 774
 QY 770 HLIHSR-LHSGEKYQCDKCGKRFSGSHSGSYQHNNHRSYCKRAEAREAREAREK 828
 Db 775 NLQRRVRNHNKPKFKCHLCYRCFGQQTNLDRHL- - - - -KKHENG 815
 QY 829 HLGPTELLMNRAYLQSIITPQGYSDSEERSEMPDESEKEHEKE- - -GEGYGLRRDG 885
 Db 816 NMSGI- - - - -ATSSPHSELESTCAIIDDKEDAVFTIRNFIGNSHSGSPRV 864
 QY 886 DE- - - - -EEEEEESEENKSM- - -DTDPETIRDEEFTGDHSDSDGCKMETKSDH 934
 Db 865 EERNMGSHFEKKALVPSQNSDLDDDEVEDEVLLDER- - -DEDYDITGKTKEPEVTSNL 921
 QY 935 EEDNMD 941
 Db 922 HEGNPD 928

RESULT 13

A31591

transcription regulator Evi-1 - mouse

C:Species: Mus musculus (house mouse)

C>Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 02-Jul-1998

C:Accession: A31591

R:Morishita, K.; Parker, D.S.; Mucenaki, M.L.; Jenkins, N.A.; Copeland, N.G.; Ihle, J.N.

Cell 54, 831-840, 1988

A>Title: Retroviral activation of a novel gene encoding a zinc finger protein in IL-3-de

A:Reference number: A31591; MUID:88311086; PMID:2842066

A:Accession: A31591

A:Molecule type: mRNA

A:Residues: 1-1042 <MOR>

A:Cross-references: GB:M21829

C:Keywords: DNA binding; zinc finger

Query Match 6.9%; Score 345; DB 2; Length 1042;
 Best Local Similarity 19.9%; Pred. No. 2.5e-10;
 Matches 212; Conservative 140; Mismatches 367; Indels 344; Gaps 47;

QY 9 RKFCTEGCAKFKYKHLKHLRHSGEKPYECPCNCKKRFSGSYSHISKKCIGLS 68
 Db 101 REYKCDQCPKAFNMKSNLIRHQMSHDSGKHVECCNCAKYFTDPSNLQRIRSGHV- - - 155
 QY 69 VNGRRNNIKTCSSPNSVSSPTNSAITQLRNKLENGKPL- - -SMSBOTGLIKIT 121
 Db 156 - - - - -GARAHACPCGCKTFATISSGLKQKHHSVKEPFCEVCHKSYTQFSNLCRHKR 208
 QY 122 EPLD- - - - -FNDYKVLMTATHGFSG- - -SSPFM- - - 145
 Db 209 MHADERTQIKCQDCQMFSTTSLNKHRRFCGKKNHFAAGCFPGQISLIPGPAMDKISM 268
 QY 146 - - - - -NGCL- - - - -GATSP- - - - -GVHPSAQSPOHLGVGEAPLIGF 179
 Db 269 VMMSHANPGLADYFGTNRHPAGLTFPTAFGFSFGLPPS- - -GLVHRPLPIPA 320
 QY 160 PTMNSNLISVQKVLQVDTVSRQKMDCKTEDISLKGVMMDPCSQPREQGVTSNIP 239
 Db 321 SPVPKGLSTEQSNKQSPLLTHPOLLPATODILKALSGH- - - - -PP 362
 QY 240 VG- - - - -LPVSHNGATKSIIDYTLKAYNE- - -AKACLOSITDSRQISNKKKELRT 290
 Db 363 VQDNKPVELLP- - - - -ERSSEERPLEKISDQSSSDLDVDTSPGSLDFTSSGDLAS 415
 QY 231 LIDLVTDDKMIENHISITPFSQCFKESFPGLPLHGHERRYCKMNEIKAVLQPHENT- 349
 Db 416 - - - - -DLSDDKEX- - - - -CKEN- - - - -GKMFKDYKSPLONA 442
 QY 350 -VPNKAGVVDNKALLSLVSEKGLTSPIN- - -PYQDMSVLKAYYAMNEPESDELKIS 407
 Db 443 STTNKKE- - -HNHSHVSASVEQSAVSAVGVNDSTIKAIASIAEKYF- - -GSTGLVGIQ 494

QY 408 I- - - - -AVGLPOEFYKWEF- - -QRKYQYNSRSRSLERTSKPLADNSNPTTKDLSLPS 461
 Db 495 DKVYCALPYPMFLPFPFAPFSQSWYF- - - - -PDRDRLSLPL- - - - -KMEPQSP 539
 QY 462 VKPMDSITSPSIAELHNSVTSCDPLRLTKSSHFNINAVDKLHSRNTSPNLST 521
 Db 540 - - - - -SEVKLQKGSSESPDLTYTKRD 562
 QY 522 SKNSHSSYTPNSFSEELQAEPLDLSLPKQREBPCKIATNKTKATSINDHNSVSS 581
 Db 563 EKPLTSGSKSGTAT- - -QDQPLDLSGSKR- - - - -ASGKLTLEPRNNHYFGEXK 613
 QY 562 SENS-EPNLITFKKEFSNNSNLDNKSNNPVFGNNP- - -SAKELYPLPQSAFPAT 637
 Db 614 GSNMOTRP- - - - -SSDQSLQHAPTPPF- - -MDPIYRVEKRLTLPLEALKE- - - 657
 QY 638 FMPPVQTSIPGLRPYRGL- - -DQMSFLPHMAYTYTGATPADMQRRKYQKQFQDL 694
 Db 658 - - - - -KYLRPSPGLFHPQMSAIENMA- - - - -EKLESFSA-L 688
 QY 695 LDGAQDYMSGLDDMTD- - -SDCLSRKKIKKTESGMVACDLCKTFQKSSSLRHKTEHT 751
 Db 689 KPEASELLQSVPSMSPFAPRNTLPENLIRKQKE-RYTCRYCGKIFPRSANLTRHLRHT 747
 QY 732 GKRPHQCOICCKAFKHKHLIHSR-LHSGEKPYQCDKCGKRFSGSYSQHNNHRSYC 810
 Db 748 GQOPFRCKYCDRFSFISNLRHVANNHNKEKPFCHLCDCRCFGQQTNLDRHL- - - 800
 QY 811 KRAEERAEAREAKHGLPTELLMNRAYLQSI- - -TPQY- - - - -SDSEER 856
 Db 801 - - - - -KKHENGMSGTATSPHSELESGAIIIDKEDAVFTIRNFIGNSHG 848
 QY 857 ESMRDEGESEK- - -BHEKEGEGYGLRRD- - -GDEEESE- - -EESEENKMDTPE- - - 906
 Db 849 SOSPRNMEERNMGSIKPKD-KKALATSQNSDLDDDEVEDEVLLDEDEDNDIPGRKEL 907
 QY 907 -TIRDEETGDHSDSDGCKMETKSD- - -HEEDNMDQM 943
 Db 908 GVTIRDEELPE- - -DDYEEAGLEMSCKASPVRYVEEDYKSL 947

RESULT 14

T14757

hypothetical protein DKFZp572C163.1 - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 01-Dec-2000

C:Accession: T14757

R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, August 1999

A:Reference number: Z18181

A:Accession: T14757

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-701 <MAN>

A:Cross-references: EMBL:AL110217

A:Experimental source: adult subthalamic nucleus; clone DKFZp572C163

C:Genetics:

A:Note: DKFZp572C163.1

C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
 Query Match 6.6%; Score 328.5; DB 2; Length 701;
 Best Local Similarity 19.7%; Pred. No. 1e-09;
 Matches 162; Conservative 87; Mismatches 251; Indels 322; Gaps 31;

QY 7 GNRKFKTECGAKFKYKHLKHLRHSGEKPYECPCNCKKRFSGSYSHISKKCIGL 66
 Db 156 GEKPYECTECGKTFKSTHLRAHQRIHTEKRYECVCECKTFSH- - -KTHLS- - - 204
 QY 67 ISVNGRMNNIKTQSSP- - -NSVSSPTNSAITQLRNKLENG-KPLMS- - -EOTGLIKIX 120
 Db 205 - - - - -VHORVHTGKPYECNDCKGSFTYNSALRAHQRIHTEKRYECSDCKT- - - 252
 QY 121 TEPLDNDYKVLMTATHG- - -SGSSPFMNGLAGATSPGLGVHPSAQSPOHLGVGEAPL- - - 176

```

Db      253 -----FAHNSALAAHRIHTEKPYECNEGCRSF---AHTSVLKAHQRIHTG-EKPYECN 303
QY      177 -----LGPPTMNSLSEYQKVLQIVDMTVSRQKMC-----KTEDISKLGKGYHMK-----D 222
Db      304 ECGRSF-TYNSALRAHQRI-----HTGRKPYECSCDECKTFAHNSALKIHQRIHTGEKPY 356
QY      223 PCSQPEEGVTSFNPVGLPVVSHNGATKSIID-YLLEKVNAAKACLOSLTDSRRQIS 281
Db      357 ECEBECK-----TFAHNSALRAHQRIHTGEKPYECSECKTF-----393
QY      282 NIKKEKRLTIDLVTDKMIENHSISPPFCQCKSPFPPIPLHGERVLCGMNEIK 341
Db      394 -PQKTLSTHRRRIHTGEK-----PYECSCGKTFQSKSYLSGHER-----IHT 435
QY      342 VLQPHENIYPNKAGVFQDNKALLLSVL--SEKGLTSPINPYKHMSVLKAYYAMNEPN 399
Db      436 GEKPYECNCGK--TFYVYKALIVHQRIHTGEK-----PYECNQ-----472
QY      400 SDELKISIAVGLPQEFVKEMFEQKRYQYNSRSRPSLEKTSKPLAPNSPTTKDILLPR 459
Db      473 -----CGKTFQQR--THLCAHQRIHTGEKPYECNECKGT-----504
QY      460 SPYKPMDSITSPSIAELHNSVTSCDPLRLTKSSHFTNIAVADKLDHSRNTSPILNLS 519
Db      505 -----FADNSALRAHRIHTGEKPYECNDCG-----KTFSTKSHLR-----HLRT 545
QY      520 TSSKNSHSSSYTNSFSSEELQAEPLDLSLPKOMREPKGIATKNTKATSIINDHNSVS 579
Db      546 RSEKPYECSECKTSEK-----SYVSAH 570
QY      580 SSENDEPLNLTFFIKKFSNSNLDKNSNPVFGMNPFSAKPLYPPLPQSAFPATPM 639
Db      571 QRVHTKEKPYECNVCCKPFAHNSTL-----595
QY      640 PAVQTSIPGLRPFGLDQMSFLPHMAVTPYTGAAATPADMQORRKYQKQSGDLDLDAQ 659
Db      596 -----RVHQRH-----602
QY      700 DYMSGLDMDTSDSCLSRKIKKTESGMVACDLCDTFOKSSILRHKYEHTGKRPPOCQ 759
Db      603 -----TGEKPYECNDCGKTFQSKSHLSAQRIHTGEKPYECN 639
QY      760 ICKKAFGKHHLIEHSRLSHGKPYQCKCKGKPSHSGSYQH 802
Db      640 ECGKAFQNSLTHRVHQRIHTGEKPYECDECKTFVRKALRVH 682

RESULT 15
S26823
zinc finger protein ZNF43 - human
N:Alternate names: zinc finger protein kox27
C:Species: Homo sapiens (man)
C:Date: 13-Jan-1995 #sequence revision 13-Jan-1995 #text_change 01-Dec-2000
C:Accession: S26823; I37967; S10416
R:Overlaping, R.; Trowsdale, J.
Nucleic Acids Res. 19, 2921-2928, 1991
A:Title: A gene encoding 22 highly related zinc fingers is expressed in lymphoid cell li
A:Reference number: S26823; MUID:91279444; PMID:1711675
A:Accession: S26823
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-803 <LOV>
R:Residues: 1-803 <LOV>
A:Cross-references: EMBL:X59244; NID:G38031; PIDN:CAA41932.1; PID:G38032
New Biol. 2, 363-374, 1990
A:Title: Multiple genes encoding zinc finger domains are expressed in human T cells.
A:Reference number: I37949; MUID:91145339; PMID:2288909
A:Accession: I37967
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 476-531 <THI>
A:Cross-references: EMBL:X52358; NID:G34160; PIDN:CAA36584.1; PID:G930090

```

C:Genetics:
A:Gene: GDB:ZNF43; HTF6
A:Cross-references: GDB:128653
A:Map position: 19p13.1-19p12
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: DNA binding; zinc finger

Query Match .66%; Score 327; DB 2; Length 803;
Best Local Similarity 19.0%; Pred. No. 1.5e-09;
Matches 160; Conservative 86; Mismatches 196; Indels 402; Gaps 31;

```

QY      7 GNRKPYECGCAFKKHLKXELRHSGEKPYECNCKKRFSGSYSHSSKKICIGL 66
Db      276 GEKPYECGCAFAFNQSSNLTKHKLHGEKPYECGCAFAFNQSSNLTKH 326
QY      67 ISVGNRRNNIKTSSPNSVSSPTNSAITOLRNKLENGKPLSMSEOTGLIKTEPLDF 126
Db      327 -----KRHTGEKPYECGCAFAFNQSSNLTKHKLHGEKPYECGCAFAFNQSSNLTKH 363
QY      127 NDYKVLMAHGFSGSPFNNGGLGATSPGLVHPSAOSPMQHLGVGMEADLGLPPTNSNL 186
Db      364 --YKTECEGEARSRS-----SNL 379
QY      187 SEYQKVLQIVDMTVSRQKMDKTEDI-----SKLGKGYHMKPCSQP---EEQVTSFN 236
Db      380 TKHKK-----IHTGEKPYECGCAFAFNQSSNLTKHKLHGEKPYECGCAFAFNQSSNLTKH 430
QY      237 IPPVGLPVVSHNGATKSIIDYTLKVNAAKACLOSLTDSRRQISNIKEKRLTIDLV 296
Db      431 WPST--LTKHRI-----HTGEKPYECGCAFAFNQSSNLTKHKLHGEKPYECGCAFAFNQSSNLTKH 464
QY      297 DDMENHSISPPFCQCKSPFPPIPLHGER-----YLCKMNEIKAVLQPHENI 349
Db      465 THKRI--HTAEPKYECGCAFAFNQSSNLTKHKLHGEKPYECGCAFAFNQSSNLTKH 509
QY      350 VPKKAGVFPDNKALLLSVLSKGLT-SPINRY-----DHMSVLKAYYAMNEPN 400
Db      510 -----GKAFKSSKLTHTKTHTEKPYECGCAFAFNQSSNLTKHKLHGEKPYECGCAFAFNQSSNLTKH 553
QY      401 DELKISIAVGLPQEFVKEMFEQKRYQYNSRSRPSLEKTSKPLAPNSPTTKDILLPR 460
Db      554 -----HTGEKPYECGCAFAFNQSSNLTKHKLHGEKPYECGCAFAFNQSSNLTKH 577
QY      461 PVPKMDSTSPSIAELHNSVTSCDPLRLTKSSHFTNIAVADKLDHSRNTSPILNLS 520
Db      578 -----577
QY      521 SSKNSHSSSYTPNSF--SSRELQAEPLDLSLPKOMREPKGIATKNTKATSIINDHNSVS 579
Db      578 -----HKIHTGEKPYECGCAFAFNQSSNLTKHKLHGEKPYECGCAFAFNQSSNLTKH 592
QY      580 SSENDEPLNLTFFIKKFSNSNLDKNSNPVFGMNPFSAKPLYPPLPQSAFPATPM 639
Db      593 -----CKATQSSNLT-----THKKIH-----611
QY      640 PAVQTSIPGLRP-----PGLDQMSFLPHMAVTPYTGAAATPADMQORRKYQKQSGDLDLDAQ 690
Db      612 -----GKPYECGCAFAFNQSSNLTKHKLHGEKPYECGCAFAFNQSSNLTKH 653
QY      691 QGDLADQADYMSGLDMDTSDSCLSRKIKKTESGMVACDLCDTFOKSSILRHKYEHT 750
Db      654 KW-----SSTLTNKHITHTGEKPYECGCAFAFNQSSNLTKHKLHGEKPYECGCAFAFNQSSNLTKH 694
QY      751 TGRPHQCOICKKAFGKHHLIEHSRLSHGKPYQCKCKGKPSHSGSYSHQHN-HRYG 809
Db      695 TGEKPYECGCAFAFNQSSNLTKHKLHGEKPYECGCAFAFNQSSNLTKHKLHGEKPYECGCAFAFNQSSNLTKH 750
QY      810 CKK 813
Db      751 TKKQ 754

```

Job time : 35 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2004, 16:23:23 ; Search time 17 Seconds

(without alignments)
2611.365 Million cell updates/sec

Title: US-09-964-238-2

Perfect score: 4991
Sequence: 1 MLTQAGNRKFKTECGRAF.....DGKMKTSDEHEDNEDMG 944

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	ID	Description
1	4891	100.0	1215	1	SIP1_MOUSE	Q9t0g7 mus musculus
2	4869	97.6	1214	1	SIP1_HUMAN	Q60315 homo sapien
3	1665	33.4	1114	1	TCF8_CHICK	P36197 gallus gall
4	1657.5	33.2	1124	1	TCF8_HUMAN	P37275 homo sapien
5	1599	33.0	1043	1	TCF8_MESAT	Q60542 mesocricetu
6	1587.5	31.8	1117	1	TCF8_MOUSE	Q64318 mus musculus
7	1550.5	31.1	1109	1	TCF8_RAT	Q62947 rattus norv
8	819.5	16.4	1060	1	ZFHL_DROME	P28166 drosophila
9	407	8.2	1845	1	ZFHL_HUMAN	Q9u136 homo sapien
10	379	7.6	1191	1	EV11_HUMAN	Q05481 homo sapien
11	355.5	7.1	1051	1	EV11_HUMAN	Q03112 homo sapien
12	354.5	7.1	1350	1	XFIN_XENLA	P08045 xenopus lae
13	345	6.9	1042	1	EV11_MOUSE	P14404 mus musculus
14	341.5	6.8	1167	1	Z208_HUMAN	Q43345 homo sapien
15	339.5	6.8	783	1	ZF25_HUMAN	Q9u115 homo sapien
16	336	6.7	1276	1	PRD3_HUMAN	Q9haz2 homo sapien
17	331.5	6.6	751	1	Z184_HUMAN	Q99676 homo sapien
18	327	6.6	803	1	ZN43_HUMAN	P17038 homo sapien
19	315.5	6.3	469	1	Z115_HUMAN	P52742 homo sapien
20	310.5	6.2	616	1	ZN33_HUMAN	P35789 homo sapien
21	297	6.0	675	1	ZG20_XENLA	P18714 xenopus lae
22	296	5.9	580	1	ZF35_MOUSE	P15620 mus musculus
23	293	5.9	818	1	KR18_HUMAN	Q9hcg1 homo sapien
24	292	5.9	738	1	ZN84_HUMAN	P51523 homo sapien
25	290	5.8	1300	1	SAL3_HUMAN	Q9bxa9 homo sapien
26	289.5	5.8	898	1	Z071_XENLA	P18751 xenopus lae
27	289.5	5.8	1323	1	SAL3_MOUSE	Q62255 mus musculus
28	287.5	5.8	947	1	Z268_HUMAN	Q14587 homo sapien
29	287.5	5.8	1053	1	SAL4_HUMAN	Q9u1q4 homo sapien
30	285.5	5.7	636	1	ZF90_MOUSE	Q61367 mus musculus
31	285	5.7	856	1	PRD1_MOUSE	Q60636 mus musculus
32	281	5.6	626	1	Z189_HUMAN	Q75820 homo sapien
33	280.5	5.6	511	1	EGRI_BRARE	P26632 brachydanio

34	280.5	5.6	1324	1	SAL1_HUMAN	Q9nsc2 homo sapien
35	278.5	5.6	698	1	Z234_HUMAN	Q14588 homo sapien
36	278.5	5.6	839	1	Z347_HUMAN	Q96567 homo sapien
37	277.5	5.6	780	1	Z084_XENLA	P18753 xenopus lae
38	277	5.5	594	1	ZF37_MOUSE	P17141 mus musculus
39	276	5.5	1029	1	Z197_HUMAN	Q14709 homo sapien
40	275	5.5	810	1	Z33A_HUMAN	Q05730 homo sapien
41	274.5	5.5	508	1	EGRI_RAT	P08154 rattus norv
42	274.5	5.5	789	1	PRD1_HUMAN	Q75626 homo sapien
43	274	5.5	533	1	EGRI_MOUSE	P08046 mus musculus
44	273.5	5.5	521	1	Z286_HUMAN	Q9hbt8 homo sapien
45	273.5	5.5	1322	1	SAL1_MOUSE	Q9et74 mus musculus

ALIGNMENTS

RESULT 1
ID SIP1_MOUSE STANDARD; PRT; 1215 AA.
AC Q9t0g7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Zinc finger homeobox protein 1b (Smad interacting protein 1).
GN ZFXH1B OR ZFX1B OR SIP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99329065; PubMed=10400677;
RA Vercheren K., Remacle J.E., Collart C., Kraft H., Baker B.S.,
RA Tydzanowski P., Nelles U., Wuytens G., Su M.-T., Bodmer R.,
RA Smith J.C., Huylebroeck D.,
RT "SIP1, a novel zinc finger/homeodomain repressor, interacts with Smad
proteins and binds to 5'-CACCT sequences in candidate target genes";
RT J. Biol. Chem. 274:20489-20498(1999).
RL
CC - FUNCTION: TRANSCRIPTIONAL INHIBITOR THAT BINDS TO DNA SEQUENCE 5'-
CACCT-3' IN DIFFERENT PROMOTERS.
CC - SUBUNIT: BINDS ACTIVATED SMAD1, ACTIVATED SMAD2 AND ACTIVATED
CC SMAD3; BINDING WITH SMAD4 IS NOT DETECTED.
CC - SUBCELLULAR LOCATION: Nuclear.
CC - SIMILARITY: BELONGS TO DELTA-BF1/ZFH-1 FAMILY OF TWO-HANDED ZINC
CC FINGER/HOMEOBOXIN PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF033116; AAD56590.1; -
CC TRANSFAC: T04864; -
DR MGD; MGI:1344407; Zfx1b.
DR GO; GO:0005634; C:nucleus; IC.
DR GO; GO:0019208; F:phosphatase regulator activity; ISS.
DR GO; GO:0016564; F:transcriptional repressor activity; NAS.
DR GO; GO:0016481; F:negative regulation of transcription; IC.
DR GO; GO:0007399; P:neurogenesis; ISS.
DR GO; GO:0007183; P:SMAD protein heteromerization; ISS.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR007087; Znf C2H2.
DR Pfam; PF00096; zF-C2H2; 8.
DR Prodom; PD000003; Znf C2H2; 1.
DR SMART; SM00389; HOX; 1.
DR SMART; SM00355; Znf C2H2; 8.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 6.
KW Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;

KM	Homeobox; Repressor; Metal-Binding, Repeat.
FT	DOMAIN
FT	ZN_FING
FT	ZN_FING
FT	ZN_FING
FT	ZN_FING
FT	DNA_BIND
PT	ZN_FING
PT	ZN_FING
FT	ZN_FING
FT	DOMAIN
SQ	SEQUENCE

Query Match	100.0%	Score 4991	DB 1	Length 1215
Best Local Similarity	100.0%	Pred. No. 1	9e-229	
Matches 94	Conservative 0	Mismatches 0	Indels 0	Gaps 0

Qy	1	MLTQAGRRKRCCTECGAFKFKHLLKXHLRJHSEKPYECPCNKKRFSHSSYSHHSS	60
Db	272	MLTQAGRRKRCCTECGAFKFKHLLKXHLRJHSEKPYECPCNKKRFSHSSYSHHSS	331
Qy	61	KKICIGLISVNGRMNRNIIKTGSSPNSSVSSPTNSAITQJLRNKLKNGKPLMSBQJLKLK	120
Db	332	KKICIGLISVNGRMNRNIIKTGSSPNSSVSSPTNSAITQJLRNKLKNGKPLMSBQJLKLK	391
Qy	121	TEPILDFNDYKYLMAHGFSSGSPFPNNGLGARSPLGYNPBAOSPMOHLGVGHEAPLGF	180
Db	392	TEPILDFNDYKYLMAHGFSSGSPFPNNGLGARSPLGYNPBAOSPMOHLGVGHEAPLGF	451
Qy	181	TMSNLSSEVQKYLQIVDNTVSRQKMDCTEDISKLKGHYMKPCSQPBEQGYTSNIPV	240
Db	452	TMSNLSSEVQKYLQIVDNTVSRQKMDCTEDISKLKGHYMKPCSQPBEQGYTSNIPV	511
Qy	241	GLPVVSHNGATKTSIIDYTLLEKYNKACLOSLITDSRROIISNKKKRLTILDTVDDK	300
Db	512	GLPVVSHNGATKTSIIDYTLLEKYNKACLOSLITDSRROIISNKKKRLTILDTVDDK	571
Qy	301	IENHSISIPSCQFCKESFPGBPIPLHONERYLCKNNBEIKALQJHENIVPKAGVFDN	360
Db	572	IENHSISIPSCQFCKESFPGBPIPLHONERYLCKNNBEIKALQJHENIVPKAGVFDN	631
Qy	361	KALLSSVLSSEKGLTSPINPYKDHSVLYKAYYAMNMEPNSEDLKISIAVGJPOEFVXEM	420
Db	632	KALLSSVLSSEKGLTSPINPYKDHSVLYKAYYAMNMEPNSEDLKISIAVGJPOEFVXEM	691
Qy	421	FEQRKVYQYSNRSRSLERTSKPLPNSNPPTKDSILPRSPKPMDSITSPSIAELHNSV	480
Db	692	FEQRKVYQYSNRSRSLERTSKPLPNSNPPTKDSILPRSPKPMDSITSPSIAELHNSV	751
Qy	481	TSCDPLRLTSSHFTNIIKAVDKLDHSRNPSPPLNLSSTSKNSHSSSYTNSFSSEEL	540
Db	752	TSCDPLRLTSSHFTNIIKAVDKLDHSRNPSPPLNLSSTSKNSHSSSYTNSFSSEEL	811
Qy	541	QAEPLDLSLPKQMRBPKGIATKNTKTKATSINLDHNSVSSSSENSEDEPLNTFIKKEFSN	600
Db	812	QAEPLDLSLPKQMRBPKGIATKNTKTKATSINLDHNSVSSSSENSEDEPLNTFIKKEFSN	871
Qy	601	SNNLDNKNNVFGNPNPSAKPLYPLRPOSAFPAPATMPPVQTSIIGLRPRPGJDOMSF	660
Db	872	SNNLDNKNNVFGNPNPSAKPLYPLRPOSAFPAPATMPPVQTSIIGLRPRPGJDOMSF	931
Qy	661	LPHMAVYTPTGATFADMQQRKRYORQKGFQDLDLDGADYVSGJDDMTDSSCSLRKTI	720
Db	932	LPHMAVYTPTGATFADMQQRKRYORQKGFQDLDLDGADYVSGJDDMTDSSCSLRKTI	991
Qy	721	KKTBSGMYACDLCDKTFQKSSLLRHXYEHTGKRPQCOICCKAFYKHHLIEHSRLHSG	780
Db	992	KKTBSGMYACDLCDKTFQKSSLLRHXYEHTGKRPQCOICCKAFYKHHLIEHSRLHSG	1051
Qy	781	EKPYOCDGCGRFSHSGSYSOHMMNRYSCKREAEERAEAEAEAEKGLGTELLMNA	840
Db	1052	EKPYOCDGCGRFSHSGSYSOHMMNRYSCKREAEERAEAEAEAEKGLGTELLMNA	1111

Qy	841	YLOSITIQGVSDSEERSMPPDGSSEKHEKEGEGYCKLRRPDDEEESESEENKS	900
Db	1112	YLOSITIQGVSDSEERSMPPDGSSEKHEKEGEGYCKLRRPDDEEESESEENKS	1177
Qy	901	MDMDPETIRDEEFTGDHSMDSSEDKKETSQSDHEEDNMEEGMG	944
Db	1172	MDMDPETIRDEEFTGDHSMDSSEDKKETSQSDHEEDNMEEGMG	1215

RESULT 2	
SIP1_HUMAN	
ID_SIP1_HUMAN	STANDARD;
Accession	PRT; 1214 AA

DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 15-SEP-2003 (rel. 42, Last annotation update)
DE Zinc finger homeobox protein 1b (Smad interacting protein 1)
DE (SMADIP1) (HRIHFB2411).
GN ZFX1B OR ZFX1B OR KIA0569.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

RX MEDLINE=98290545; PubMed=9628881;
 RA Nagaee T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
 RA Nomura N., Ohata O.,
 RI "Prediction of the coding sequences of unidentified human genes. IX
 RI The complete sequences of 100 new cDNA clones from brain which can
 RI code for large proteins in vitro.";
 RI DNA Res. 5:31-39(1998).

R6
R7
R8 SEQUENCE FROM N.A., AND DISEASE.
R9
RA Pubmed=11448942;
RB Cacheux V., Dastot-Le Moal F., Keeseleirenen H., Bondurand N.,
RC Rostalet A., Botsister B., Wilson M., Mowat D., Goossens M.;
RD "Loss-of-function mutations in STPL Smad interacting protein 1 result
RE in a syndromic Hirschsprung disease." ;
RF Hum. Mol. Genet. 10:1503-1510(2001).
RG

RN 13)
 RP SEQUENCE FROM N.A.
 RA MEDLINE=21157348; PubMed=11279515;
 RX Wakamatsu N., Yamada Y., Yamada K., Ono T., Nomura N., Taniguchi H.,
 RA Kitoh H., Mutoh N., Yamanaoka T., Mushiike K., Kato K., Sonta S.,
 RA Nagaya M.;
 RT "Mutations in SIP1, encoding Smad interacting protein 1, cause a form
 of Hirschsprung disease.";

RL Nat. Genet. 27:369-370 (2001).

RN [4]

RP SEQUENCE OF 1161-1214 FROM N.A., AND SUBCELLULAR LOCATION.

RC TISSUE=Petal brain;

RX MEDLINE=99068504; PubMed=9853515;

RA Ueki N., Oda T., Kondo M., Yano K., Noguchi T., Muramatsu M.A.;

RT "Selection system for genes encoding nuclear-targeted proteins";

RL Nat. Biotechnol. 16:1133-1142(1998)

CC -1- FUNCTION: TRANSCRIPTIONAL INHIBITOR THAT BINDS TO DNA SEQUENCE
CC -1- "CACC'-3'" IN DIFFERENT PROMOTERS (BY SIMILARITY).
CC -1- SUBUNIT: BINDS ACTIVATED SMAD1, ACTIVATED SMAD2 AND ACTIVATED
CC -1- SMAD3; BINDING WITH SMAD4 IS NOT DETECTED (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DISEASE: Defects in ZFPX1 are a cause of Hirschsprung disease
CC associated with microcephaly, mental retardation, hypertelorism,
CC submons cleft palate and short stature.

CC 1-1 SIMILARITY, BELONGS TO DELTA-EPI/ZEPI-1 FAMILY OF TWO-HANDED ZINC
CC PINNER/HOMODOMAIN PROTEINS.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collabora-
CC tion between the Swiss Institute of Bioinformatics and the EMBL outposts
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in the public
CC domain.

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AB01141; BAA25495.1; -
 CC EMBL; AY029472; AAK52081.1; -
 DR EMBL; AB056507; BAB40819.1; -
 DR EMBL; AB015341; BAA34798.1; -
 DR TRANSFAC; T05057; -
 DR Genew; HGNC:14881; ZFXH1B.
 DR MIM; 605802; -
 DR MIM; 235730; -
 DR GO; GO:0005634; C:nucleus; IC.
 DR GO; GO:0019208; F:phosphatase regulator activity; NAS.
 DR GO; GO:0046332; F:SMAD binding activity; NAS.
 DR GO; GO:0016564; F:transcriptional repressor activity; ISS.
 DR GO; GO:0016481; P:negative regulation of transcription; IC.
 DR GO; GO:0007399; P:neurogenesis; NAS.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR007087; Znf C2H2.
 DR Pfam; PF00096; zf-C2H2; 8.
 DR Prodom; PD000003; Znf C2H2; 1.
 DR SMART; SM00389; HOX; 1.
 DR SMART; SM00355; Znf C2H2; 8.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 6.
 KW Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger; Homeobox; Repressor; Metal-binding; Repeat; Hirschsprung disease.
 FT DOMAIN 437 487
 FT ZN_FING 211 234 C2H2-TYPE.
 FT ZN_FING 241 263 C2H2-TYPE.
 FT ZN_FING 282 304 C2H2-TYPE.
 FT ZN_FING 310 334 C2H2-TYPE (ATYPICAL).
 FT DNA_BIND 644 703 HOMEBOX-LIKE.
 FT ZN_FING 999 1021 C2H2-TYPE.
 FT ZN_FING 1027 1049 C2H2-TYPE.
 FT ZN_FING 1055 1076 C2H2-TYPE (ATYPICAL).
 FT ZN_FING 1084 1214 GLU-RICH (ACIDIC).
 FT DOMAIN 1084 1214
 SO SEQUENCE 1214 AA; 136446 MW; B578FD91339C3FDD CRC64;

Query Match 97.6%; Score 4869; DB 1; Length 1214;
 Best Local Similarity 97.5%; Pred. No. 1.2e-223;
 Matches 919; Conservative 13; Mismatches 11; Indels 0; Gaps 0;

1 MVTGAGRRKFKCTCGAKFKYKHLKEHLRHSGEKPEPCNCKKRRFSHSGSYSHISS 60
 272 MVTGAGRRKFKCTCGAKFKYKHLKEHLRHSGEKPEPCNCKKRRFSHSGSYSHISS 331
 61 KKCIGLISVNGMRNRNITKGSPPNSVSSPTNSAITOLRNKLENGKPLMSBQTLKIX 120
 332 KKCIGLISVNGMRNRNITKGSPPNSVSSPTNSAITOLRNKLENGKPLMSBQTLKIX 391
 121 TEPLDPRNYKVLMAHFGSGSPFNWGLGATSPGVHSPASPMOHLGVGHEALVLP 180
 392 TEPLDPRNYKVLMAHFGSGSPFNWGLGATSPGVHSPASPMOHLGVGHEALVLP 451
 181 TMSNLSSEVOYKVLQIVDNTVSRQMDCKTEDISKLGKYMPCGSOPEOQVTSNIPV 240
 452 TMSNLSSEVOYKVLQIVDNTVSRQMDCKTEDISKLGKYMPCGSOPEOQVTSNIPV 511
 241 GLPVVSHNGATKSIIDYTLKYNKAKALQSLTTDSRRQISNIKKKELRTLLDVTDDKM 300
 512 GLPVVSHNGATKSIIDYTLKYNKAKALQSLTTDSRRQISNIKKKELRTLLDVTDDKM 571
 301 IENHSISIPFSQCFCKESPPGPIPLHGHRRYKCKANNEIKAVLQPHENIVPKKAVFPDN 360
 572 IENHSISIPFSQCFCKESPPGPIPLHGHRRYKCKANNEIKAVLQPHENIVPKKAVFPDN 631
 361 KALLSSVLSSEKGLSPINPYKDHMSVLYKAYAMMEPNSDELKISTAVGLPOEFVKEM 420
 632 KALLSSVLSSEKGLSPINPYKDHMSVLYKAYAMMEPNSDELKISTAVGLPOEFVKEM 691
 421 FEQRKYVOYSNRSPLERTSKPLAPNSNPPTKDSLLPRSPVYPKMDSITSPSIABLHNSV 480

692 FEQRKYVOYSNRSPLERTSKPLAPNSNPPTKDSLLPRSPVYPKMDSITSPSIABLHNSV 751
 481 TSCEPPLRLTKSSHPTNIN KAVDKLDHRSNTPSPDLNLSSTSSKNSHSSSYTPNFSSEEL 540
 752 TNCDEPLRLTKSPHFTNIN KPEKLDHRSNTPSPDLNLSSTSSKNSHSSSYTPNFSSEEL 811
 541 QAEPLDLSLPKQMBEKKIATKTKYKATSINLDHNSVSSSSSENSDEPLNTFTIKKERN 600
 812 QAEPLDLSLPKQMBEKKIATKTKYKATSINLDHNSVSSSSSENSDEPLNTFTIKKERN 871
 601 SNNLDKNSNNVPGMNPSPAPLTYPLPPOGAFPPATMPVQVNSIPCLRPYPLGDMQSF 660
 872 SNNLDKNSNNVPGMNPSPAPLTYPLPPOGAFPPATMPVQVNSIPCLRPYPLGDMQSF 931
 661 LPHAAVYPTGAATFADMOQRKRYKQKQFGCDLLDGAODVMSGLDMDTSDSCSRKXI 720
 932 LPHAAVYPTGAATFADMOQRKRYKQKQFGCDLLDGAODVMSGLDMDTSDSCSRKXI 991
 721 KTESGMYACDLCDTFOKSSLLRHKYETKRPKPOCICKKAPKHKHLIERSRLHSG 780
 992 KTESGMYACDLCDTFOKSSLLRHKYETKRPKPOCICKKAPKHKHLIERSRLHSG 1051
 781 EKPQCDKCGRFSHSGSYSGHMMHRSYCKREAEEREAAREAREKCHLPTTEILMRA 840
 1052 EKPQCDKCGRFSHSGSYSGHMMHRSYCKREAEEREAAREAREKCHLPTTEILMRA 1111
 841 YLGSTTPQGYSDSEERESMPDGESEKHEKEGEGYGLKRRQDDEEESESESEKNS 900
 1112 YLGSTTPQGYSDSEERESMPDGESEKHEKEGEGYGLKRRQDDEEESESESEKNS 1171
 901 MDTPETIRDEEETGDHSMDDSSBDGKMETKSDHEEDMEDGM 943
 1172 MDTPETIRDEEETGDHSMDDSSBDGKMETKSDHEEDMEDGM 1214

RESULT 3
 TCF8 CHICK STANDARD; PRT; 1114 AA.
 AC P36197; 042408;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Transcription factor 8 (Delta-crystallin enhancer binding factor) (Delta E8).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lens;
 RX MEDLINE=94116444; PubMed=7904558;
 RA Funahashi J.-I., Sekido R., Murai K., Kamachi Y., Kondoh H.;
 RT "Delta-crystallin enhancer binding protein delta E8 is a zinc finger-homeodomain protein implicated in postgastrulation embryogenesis";
 RT Development 119:433-446(1993).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Embryo;
 RX MEDLINE=97082972; PubMed=8964504;
 RA Sekido R., Takagi T., Okanami M., Moribe H., Yamamura M., Higashi Y., Kondoh H.;
 RT "Organization of the gene encoding transcriptional repressor deltaE8 and cross-species conservation of its domains";
 RL Gene 173:227-232(1996).
 CC - FUNCTION: BINDS TO DELTA 1-CRYSTALLIN ENHANCER CORE AND REPRESSORS LENS-SPECIFIC TRANSCRIPTION. IT BINDS AS WELL MANY OTHER NON-LENS SPECIFIC DNA SEQUENCES.
 CC - SUBUNIT: MAY INTERACT WITH CTBP1 (BY SIMILARITY).
 CC - TISSUE SPECIFICITY: EXPRESSION IS DEVELOPMENTALLY REGULATED WITH

Science 254:1791-1794(1991).
- FUNCTION: INHIBITS INTERLEUKIN-2 (IL-2) GENE EXPRESSION. MAY BE RESPONSIBLE FOR TRANSCRIPTIONAL REPRESSION OF THE IL-2 GENE. ENHANCES OR REPASSES THE PROMOTER ACTIVITY OF THE ATPIAL GENE DEPENDING ON THE QUANTITY OF CDNA AND ON THE CELL TYPE.
- SUBCELLULAR LOCATION: Nuclear.
- TISSUE SPECIFICITY: EXPRESSED IN HEART AND SKELETAL MUSCLE, BUT NOT IN LIVER, SPLEEN, OR PANCREAS.
- SIMILARITY: BELONGS TO DELTA-BF1/ZFH-1 FAMILY OF TWO-HANDED ZINC FINGER/HOMEOBOX MAIN PROTEINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL, D15050; BAA03646.1; -.
DR EMBL, U12170; AAA20602.1; -.
DR EMBL, M81699; -. NOT_ANNOTATED_CDS.
DR PIR, JX0293; JX0293.
DR TRANSFAC, T00625; -.
DR Genew, HGNC:11642; TCF8.
DR MIM, 189909; -.
DR GO, GO:0003713; F:transcription co-activator activity; TAS.
DR GO, GO:0003714; F:transcription co-repressor activity; TAS.
DR GO, GO:0003700; F:transcription factor activity; TAS.
DR GO, GO:0008270; F:zinc ion binding activity; TAS.
DR GO, GO:0008283; P:cell proliferation; TAS.
DR GO, GO:0006955; P:immune response; TAS.
DR GO, GO:0001122; P:negative regulation of transcription from P. .; TAS.
DR InterPro, IPR01356; Homeobox.
DR InterPro, IPR07087; Znf C2H2.
DR Pfam, PF00966; Zf-C2H2; 7.
DR Prodom, PD00010; Homeobox; 1.
DR Prodom, PD000003; Znf C2H2; 2.
DR SMART, SM00389; HOX; 1.
DR SMART, SM00355; ZNF C2H2; 7.
DR PROSITE, PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE, PS0157; ZINC_FINGER_C2H2_2; 6.
KW Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger; Homeobox; Repressor; Activator; Metal-binding; Repeat.
FT ZN_FING 170 193 C2H2-TYPE.
FT ZN_FING 200 222 C2H2-TYPE.
FT ZN_FING 240 262 C2H2-TYPE.
FT ZN_FING 268 292 C2H2-TYPE (ATYPICAL).
FT DNA_BIND 581 640 HOMEBOX-LIKE.
FT ZN_FING 904 926 C2H2-TYPE.
FT ZN_FING 932 954 C2H2-TYPE.
FT ZN_FING 960 981 C2H2-TYPE (ATYPICAL).
FT DOMAIN 989 1124 GLU-RICH (ACIDIC).
FT CONFLICT 420 420 V -> I (IN REF. 2).
FT CONFLICT 609 609 E -> Q (IN REF. 3).
FT CONFLICT 654 654 I -> T (IN REF. 2).
FT CONFLICT 672 672 L -> H (IN REF. 3).
FT CONFLICT 681 681 D -> S (IN REF. 3).
SQ SEQUENCE 1124 AA; 124073 MW; 0A2714CC37C84BD1 CRC64;

Query Match 33.2%; Score 1657.5; DB 1; Length 1124;
Best Local Similarity 41.9%; Pred. No. 1.7e-71;
Matches 406; Conservative 147; Mismatches 308; Indels 107; Gaps 32;

QY 2 LTGAGNRKFKTEGKAFKTKHLEKRIHSGKPYECPCCKKRFSSGYSYSHISK 61
DB 231 VTGSCNKKFKCTEGKAFKTKHLEKRIHSGKPYECPCCKKRFSSGYSYSHISK 290
QY KCIIGLISVNGRNNIKTG--SSPNSVSSPTNSAITOLRNLKNGKPLMSSEOTGLKI 119
DB 231 KCIISLIPNGRRTGLKTSQCSPP-SLSAPSPSPRPQIRQIKEN-KPI--OEQLSVQOI 346
QY 120 KTEPLDFNDYKVLMTATGFSGSSPFMNGLGATSPLGVAHPSAQSPMOHL--GVGMEAPL 176

DB 347 KTEPVY-EFKRIYVASSINGSTPLONGVFTGGELQATSSPQGVAVLPTGLVSP 405
QY LGPFTMNSSEVQVLOI-VDNTYRSQKMDCKEDISKLGHYMKDCPSQEEGVTS 235
DB 406 -----SINLSDIQLVKAVDGNVIRQVLE---NQNANLAKSEDETINASPIQGGHVS 456
QY 236 NIPVGLPVVSHNGATYSIIDTYLEKVEAKACLOSITTDSSROISIKKEKRLTIDLV 295
DB 457 -ISALISPLVQDGTTLIIIVTSLEQSPQLGVPPNKLKKNPVATNSCKSEKLPEDLVK 515
QY 296 TD-DKMTFNH-SITPSPCOFCXKSPFQPI-PLHCHERYLCKMNEIRAVLQPHENIYN 352
DB 516 SEKKSPFGVNDSTCLCDCC--PEDINALBELKAY-----DLKQPPP-----PP 560
QY 353 KAGVFDKALLLSVSEKGLTSPINPKDMSTLKYVAMNNEPNSDELKISIAVL 412
DB 561 LPAEAKEPSSVSATDGNLSPQPLKULSLKLYALNMQPSAEELSKLADSVNL 620
QY 413 POEFVKEWFEQRYOYNSNRSPSLERTSKPLAENSPTTKDSLPRSPVPMDSITSPS 472
DB 621 PLDYVKKWFEKMGQAGQIS-----YQSESPSPPEPKV-----NIPAKNNDQPGAN 666
QY 473 IAEILNYSVSCDPPRLTKSSHFTNRYKAVDKLDHRSNTSPPLNLSSTSKSHSSYTP 532
DB 667 ANEPDSTVNLQSPDKMTNS--PVLPGSTTNGSRSSSTPSPPLNLSSSHNTQGYLYTA 723
QY 533 NSFSEELQAEPLDLSLPMKOREPKGIATKNTKATISINLDHNSVSSSSNS-----DE 587
DB 724 EG-AQEEQVPELDLSLTKQGE-----LLESTTISVYQNSVSVQEE 766
QY 588 PLNLTFFIKKESNNLNDKSNNPVFCMNPSPAKPLTLPPOSAPPPATMPPVQTSIP 647
DB 767 PLNLSCKAKEQKQSCVTD--SEPVNVNIPPSANPINAIPTVNAQLPTIYALADQNSVP 824
QY 648 GLRPYGLDQMSFLPHNAVYTPPTGAATPADMOQRKRYKQKQFGQDLIDGADIVMSGLD 707
DB 825 CLRALANKQTLLPQVAVYVTSVTP-AVOEPLKVIQPNQNDERODTSEGVSNVED 883
QY 708 MTDSDSCSRKKIKKTESGMVACLDCTPQKSSLSLHKYEHGKRPKHOICCKAKFK 767
DB 884 QNDSSTPPKKKMKRTENGMAACDLCDKIPQKSSLLHKYEHGKRPKHEGCIKAKFK 943
QY 768 KHLILHESRLSGEKPYQCDKCGKRFSSGYSQHMNRVYCKREAEREAREAREK 827
DB 944 KHLILHESRLSGEKPYQCDKCGKRFSSGYSQHMNRVYCKREAERSTGE----- 999
QY 828 GHLPTELLMRAVLIQSTP-QGYSDSSEERSMRP--DGSEKHEKEKEE-----GYGK 879
DB 1000 -EAGP-ETLSNEHVAGARASPSQG--DSDRESLTREEDDESEKEEEDKEWELQEEKE 1055
QY 880 LRRDGDDEEEERESSEKNSMDMDPETIRDEEERGHSMDSDD-----GKNET 930
DB 1056 CEKPGQDEEBEEREBEVEEBEVE-EAENEGEAKTEGLMKODRAESQASLQKVGESSE 1114
QY 931 KSDHEEDN 938
DB 1115 QVSEKTN 1122

RESULT 5
TCF8_MESAU
ID_TCF8_MESAU STANDARD; PRT; 1043 AA.
AC Q60542;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transcription factor 8 (zinc finger protein Bzp).
GN TCF8 OR BZP.
OS Mesocricetus auratus (golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.

OX NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Insulinoma;
 RX MEDLINE=95021206; PubMed=7935395;
 RA Franklin A.J., Jettion T.L., Shelton K.D., Magnuson M.A.;
 RT "B2P, a novel serum-responsive zinc finger protein that inhibits gene transcription.";
 RL Mol. Cell. Biol. 14:6773-6788(1994).
 CC -1- FUNCTION: ACTS AS A TRANSCRIPTIONAL REPRESSOR. BINDS TO E-BOX SEQUENCES IN THE IMMNOGLOBULIN HEAVY CHAIN ENHANCER AS WELL AS IN THE REGULATOR REGIONS OF MANY OTHER TISSUE-SPECIFIC GENES.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO DELTA-EFL/ZFH-1 FAMILY OF TWO-HANDED ZINC FINGER/HOMEDOMAIN PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L13856; AAC37667.1; -.
 CC PIR; A56037; A56037.
 CC InterPro; IPR001356; Homeobox.
 CC InterPro; IPR007087; Znf_C2H2.
 CC Pfam; PF00096; zf-C2H2; 7.
 CC ProDom; PD000010; Homeobox; 1.
 CC ProDom; PD000003; Znf_C2H2; 2.
 CC SMART; SM00389; HOX; 1.
 CC SMART; SM00355; Znf_C2H2; 7.
 CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
 CC PROSITE; PS0157; ZINC_FINGER_C2H2_2; 6.
 CC Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger; Homeobox; Repressor; Activator; Metal-binding; Repeat.
 CC FT ZN_FING 94 117 C2H2-TYPE.
 CC FT ZN_FING 124 146 C2H2-TYPE.
 CC FT ZN_FING 164 186 C2H2-TYPE.
 CC FT ZN_FING 192 216 C2H2-TYPE (ATYPICAL).
 CC FT DNM_BIND 504 563 HOMEBOX-LIKE.
 CC FT ZN_FING 827 849 C2H2-TYPE.
 CC FT ZN_FING 855 877 C2H2-TYPE.
 CC FT ZN_FING 883 904 C2H2-TYPE (ATYPICAL).
 CC FT DOMAIN 912 1043 GLU-RICH (ACIDIC).
 CC SEQUENCE 1043 AA; 114202 MW; 5F8BE23D1BC667 CRC64;
 SQ
 Query Match 32.0%; Score 1599; DB 1; Length 1043;
 Best Local Similarity 40.2%; Pred. No. 9.2e-69;
 Matches 390; Conservative 142; Mismatches 308; Indels 130; Gaps 30;
 QY 2 LNOGAGNRKPKCECGKAPKFKHLEKHLRIHSGEKPEKPKKPSHSGSYSHSSK 61
 DB 155 VTQSGNRRKPKCECGKAPKFKHLEKHLRIHSGEKPEKPKKPSHSGSYSHSSK 214
 QY 62 KCIGLISVNGRMNRIKGTSSPN-SVSSSPTNSAITQLANKLENGKPLMSQGTGLKIK 120
 DB 215 KCIQLMPVNGRPSGLKTCSPSLASGSPTRQIRKIKIN-KPL-QEPLSNQIK 271
 QY 121 TEPLDNDYKVLMTATGFGSSGSPFNG--GLG---ATSLGVHPSAOSPMOHLGVME 173
 DB 272 TEPLVDY-EFKPIVAVAGINCSTPLQNGVFGGQLQATSSPGQVAVVLP----VGLV 326
 QY 174 APLLGFTMNSNRSEVOKYQI-VDMTVSHQKMDCTEDISKLGKHMDCPGQPEQCV 232
 DB 327 SPT----SINSDIQLNVLKALDQAVNRQVLENNQASL----ASKQEA 368
 QY 233 TSPN----IPVGLPVSHNGATKSIIDYTLKVNKAKACLSLTDSRQISNIK 284
 DB 369 NASSIOGGSHVSAISLPLVDQDGTITIKIINSLSQPSQLQVPPQNLKIKENAPFNSCK 428
 QY 285 KEKLRTLLIDLVTDKMIENHSISTPSCQCKSPGPIPLHGHERYLCRMNEIKAVIQ 344

DB 429 SEKLPE--DLTVKSEKDKFGDAVDGCTCLLYDCCGDI-----NALPELKHVD 476
 QY 345 PHENIVPNKAGYFVNDKALLSSVLEKGLTSPIN-PYDHNHVLKAYYAMNEPNSDEL 403
 DB 477 PEHPAPPPPPAEBAEKEPSAASS--ARDGDLSPSQPLNKLILKAYYALNAPNSDEL 534
 QY 404 LKISIAVGLPOEFVKEWFEQRKRYQYSNR---SPSLERTSKPLAPNSPTTKOSLPR 459
 DB 535 SKIADSVNPLPVDVKKFKPEKMQAGIIPGQSPPPSPETGTVINPAKSDRQPADSSEQA 594
 QY 460 SPKPKMDITSSISALHNSVTSCEPLLLTYSNHTNLIKAVDKLDHSSNT--PSPLLV 517
 DB 595 E-----DSASGQS-----PLMTSS--PVLPGSALNGSSCTSPSPPLNV 633
 QY 518 SSTSKNSHSSSTPNSFSEELQAPPLDLSPKQWREPKGIATNKTATNSINDHNS 577
 DB 634 SSKRLQGY--FCVADGQOEPPQVPLDLSLPKQGE-----LLERST 674
 QY 578 VSSSSSENS-----DEPLNFTFKKESNSNNLDNKSNNPFCKNPPSAKPLTLPPOGA 632
 DB 675 ISSVQNSVYSVQEBEPLNLSVYKKEPQEDSCVTD--SEPVNVVIPSANPINALIPTVA 732
 QY 633 PPATMPPVVQTSIPGLRTPYGLDQMSFLPHMAYTTPGAATPADMQQRKTKQKQFGQ 692
 DB 733 QLETVIADQNSVPCLRALANKQTILLPOVAYTATVSP-AMQEPVVKYIQPNQND 791
 QY 693 DILDGADVYMSGLDMDTSDSCLSRKIKKTESGMVACDLCKTFQKSSSLRLHKETHG 752
 DB 792 ERQDTSSEGVSTVEDQNDSDSTPPKKTRTKNGVACDLCKTIQKSSSLRLHKETHG 851
 QY 753 KRPHQCIQKAPKFKHLEIHSRLHSGEKPYQCDKGRKFSHSGSYQHMNRYSYCKR 812
 DB 852 KRPHQCIQKAPKFKHLEIHSRLHSGEKPYQCDKGRKFSHSGSYQHMNRYSYCKR 911
 QY 813 EAEEERAAERAREKCHLPTELLMRAVLQSTTPGYSDBEREEMPPRDSSEKHEKE 872
 DB 912 EAEEERGTEQOEAEQ---LEALMNEHVGARASP-SQADSERESTLTDEEDDSKEEE 966
 QY 873 GEEGYKLLRRQDEEESESESEKSMQDTPETIRDEETGHDHSDSS--EDGKMET 930
 DB 967 EEE-----DKMEHLOEKEKCGNAQAEHEE---EEHEMDGAKDAAKTQDAVEN 1015
 QY 931 KSDHEEDNME 940
 DB 1016 GAAPQAGSLE 1025
 RESULT 6
 TCF8 MOUSE
 ID TCF8 MOUSE STANDARD; PRT; 1117 AA.
 AC 064318; 062519;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Transcription factor 8 (Zinc finger homeobox protein 1a) (MEB1) (Delta EF1).
 GN TCF8 OR ZFX1A OR ZFX1A OR ZFX1A.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Embryo;
 RX MEDLINE=97082972; PubMed=8964504;
 RA Sekido R., Takagi T., Mochi H., Yamamura M., Higashi Y., Kondoh H.;
 RT "Organization of the gene encoding transcriptional repressor deltaEFL1 and cross-species conservation of its domains.";
 RL Gene 173:227-232(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Spleen;

R	MU Y., Montoya G.D., Rubin S.E., Brodie S.G., Jenkins N.
A	Williams T.M.;
R	Submitted (May-1995) to the EMBL/GenBank/DBJ databases.
L	[3]
N	SEQUENCE FROM N.A.
P	
R	TISSUE=Brain;
R	MEDLINE=96194821; PubMed=8647466;
X	
C	"Cloning of a cDNA encoding a mouse transcriptional repressor
R	displaying striking sequence conservation across vertebrates";
RT	Gene 169:289-290(1996)".
CC	- FUNCTION: ACTS AS A TRANSCRIPTIONAL REPRESSOR. BINDS TO E-BOX
CC	SEQUENCES IN THE IMMUGLOBULIN HEAVY CHAIN ENHANCER AS WELL AS IN
CC	THE REGULATORY REGIONS OF MANY OTHER TISSUE-SPECIFIC GENES.
CC	- SUBCELLULAR LOCATION: Nuclear.
CC	- SIMILARITY: BELONGS TO DELTA-EFL/ZFH-1 FAMILY OF TWO-HANDED ZINC
CC	FINGER/HOMEODOMAIN PROTEINS.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by, and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
D	EMBL; U26259; AAA67564.1; -;
DR	EMBL; D76432; BAA1177.1; -;
DR	EMBL; L48363; AAB08442.1; -;
DR	PIR; JC4934; JCA934.
DR	TRANSFAC; T03915; -;
DR	MGI; MGI:1344313; Zfhx1a.
DR	InterPro; IPR001356; Homeobox.
DR	InterPro; IPR007087; ZnF_C2H2.
DR	Pfam; PF00096; zf-C2H2; 7.
DR	Prodorm; PD000010; Homeobox; 1.
DR	Prodorm; PD000003; ZnF_C2H2; 2.
DR	SMART; SM00389; HOX; 1.
DR	SMART; SM00355; ZnF_C2H2; 7.
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR	PROSITE; PS50157; ZINC_FINGER_C2H2_2; 6.
KM	Transcription regulation: DNA-binding; Nuclear protein; Zinc-finger;
KM	Homeobox; Repressor; Activator; Metal-binding; Repeat.
FT	ZN_FING 150 173 C2H2-TYPE.
FT	ZN_FING 180 202 C2H2-TYPE.
FT	ZN_FING 220 242 C2H2-TYPE.
FT	ZN_FING 248 272 C2H2-TYPE (ATYPICAL).
FT	DN_A_BIND 559 618 HOMEOBOX-LIKE.
FT	ZN_FING 882 904 C2H2-TYPE.
FT	ZN_FING 910 932 C2H2-TYPE.
FT	ZN_FING 938 959 C2H2-TYPE (ATYPICAL).
FT	DOMAIN 969 1117 GLU-RICH (ACIDIC).
FT	CONFICT 33 33 S->A (IN REF. 3).
FT	CONFICT 57 57 M->K (IN REF. 3).
FT	CONFICT 125 125 MISSING (IN REF. 3).
FT	CONFICT 146 146 F->L (IN REF. 3).
FT	CONFICT 313 313 E->EV (IN REF. 3).
FT	CONFICT 353 353 G->A (IN REF. 3).
FT	CONFICT 400 400 V->L (IN REF. 3).
FT	CONFICT 461 461 Q->E (IN REF. 3).
FT	CONFICT 528 528 K->KK (IN REF. 3).
FT	CONFICT 664 664 S->T (IN REF. 3).
FT	CONFICT 744 744 E->L (IN REF. 3).
FT	CONFICT 822 822 T->A (IN REF. 3).
FT	CONFICT 1062 1062 MISSING (IN REF. 3).
QO	SEQUENCE 1117 AA; 122464 MW; DIPACD2004BD34Z37 CRC64;

```

Query Match      31.8%  Score 1587.5;  DB 1;  Length 1117;
Best Local Similarity 39.9%  Pred. No. 3.5e-68;
Matches 395;  Conservative 142;  Mismatches 307;  Indels 145;  Gaps 32;

QY      2  LTGGAGNKKPCTGCGKAFKKYKHHLKELRLRIHSGEPCNCKKRFPSHSGSYSHISK 61
      :|:|||||

```

Db	211	VTOSGGRKREKCTECGAFYKXHLKHLRIHSEKREYECNCKKRSHSGSYSHHSK	270
Qy	62	KCIGLISVNGMRNNIKTY--SSEVSYSSTPNTSAITQLRNLKLENGPLSMSEQTGLKI	119
Db	271	KCISLMPNGPRRGLTKQSSSP-SLSTSPGSPTRPQIRKIEH-KPL--QEPYVNOI	326
Qy	120	KTEBLDPNDYVLMATHGFGSSPFMMNGGLGATSPLGVBAGSPMOHL--GYGMAPL	176
Db	327	KTEBVY-BEPPIVASGINSTYPLQNGVSSGGQLATSSPGQVAVVLPVGLVSP	385
Qy	177	LGPEPMNSINSEVOYVQI--VDNTVSQRKMDCKTEDIKLGYNMKPCSQPEEGYSP	235
Db	386	-----SINSDIQNLKVAVDGVIQVLETNQ--ASLASKQGEAVASAPIOQGHV	436
Qy	236	NIPVGLPVVSHNGATYSIIDTYLEKYNKAQLOSTLTSRRQISNIKKEKLTLDLV	295
Db	437	-ISASIPLVQODPTKIIINYSLQPSQLOVPQNKKELPAPTNCSKSEKLPEDLVK	495
Qy	296	TD-DKMLH-NHSITPSPCOFCKSPFG--PIPLHHERLQGMNEIKAVLOPHENIV	350
Db	496	SETDSEFGARDOSTCLCEDC---FGDLNALBELKHVDEBCAOPPPAPATEKEPSS	551
Qy	351	PNKAGVVDNKKALLSSVLSSEKGLTSPIN-PYKDMSVLKAYYAMNNEPNSDELLKISIA	409
Db	552	ASSAG-----NGDLSFSQPLKNLILKAYVALNMQSTELSKLADS	595
Qy	410	VGLPQEVKWEFEQKRYOYNSRSPS-----LETSKPLAPNSPTTKOSL	457
Db	596	VNLPLDGKKKFEQMOGOIP-QGSPPPSPGTSVNIPTKTEQOPQADNEPOEDSTR	654
Qy	458	PRSPVKMDSTSTPSIALHNSVYSCDPLRLTSSHTNIIKAVDKLHKSNTSPPLNL	517
Db	655	QGSPVKRSPVLE-VSANGNSRC-----TSBSPPLNL	688
Qy	518	SSTSSKSHSSSYTPNPSFSEELQAEPLDLSLPRKMBPKGIATKXKTKATISINLDHNS	577
Db	689	CSANPQOYS---CVAGQAGEEPVEPLDLSLPGQGE-----LLERST	729
Qy	578	VSSSENS-----DEPLNLTPIKKEFNSNUNLKNPNPFGMNPFAKLYTLPLPOSA	632
Db	730	VSSYQNSVYSVQDEPLNLSCKAKEPKQDSCVTD--SEPVNVVVPSPANPILNIPVTA	787
Qy	633	FPAPTFMPVQTSIPGLRPYVGLDQMSFLPMAVATYPLGATTFADMOOR--KYORKQGF	690
Db	788	QLPTIVAILADONSYPCLBALANKQITLIPVATY---SATVSRAVQEPVYKIQENGN	844
Qy	691	QGDLLDAQDYSGLDMDTDSCLSRKKIKTESGMYACDLCDTFOKSSSLRHKYEY	750
Db	845	QDERQDTSSEGVSTVEDQNDSDSTPPKKTKRTKENGUYACDLCXIKFOKSSSLRHKYEY	904
Qy	751	TGKRPQOQICCKAFKFKHLLIHSRLHSGRKPQCDCKGGRFBSHSGSYGQNMHRYSYC	810
Db	905	TGKRPHCIGICRKAFKFKHLLIEMRLHSGEKPYQCDCKGGRFBSHSGSYGQNMHRYSYC	964
Qy	811	KREAREEARERAREKGLQPTEL--LIMRAYQSLTPQGSYSPSERESMPR--DDESE	866
Db	965	KRGHEDDAMEGE-----DAGPVLPRVLATENHGAASPGQADSDRESITREDEDESE	1019
Qy	867	KEHEKEGEE---GYGK-LRRRDGDEEEEEESEENKSMQDTEJTIRDEEETGDH---	917
Db	1020	KEHEEHEKEMEBELOEGKECENPQGEHEEHEEHEEHEEHEEHEEVEADAEHEAATAKTGDT	1079
Qy	918	-----SMDSDSDGKMETYSIDHEE	936
Db	1080	VEVGAQAQGSLEQKASESEKESSESEQ	1108

RESULT 7			
TCF8_RAT	STANDARD;	PRT;	1109 AA
ID_TCF8_RAT			
AC	062947; 062948;		
DT	16-OCT-2001 (Rel. 40, Created)		
PT	16-OCT-2001 (Rel. 40, Last sequence update)		

DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transcription factor 8 (Zinc finger homeodomain enhancer-binding
GN protein) (zfhep).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 6-1109 FROM N.A. AND ALTERNATIVE SPLICING.
RX MEDLINE=96365389; PubMed=8769566;
RA Cabanillas A.M., Darling D.S.;
RT "Alternative splicing gives rise to two isoforms of Zfhep, a zinc
RT finger/homeodomain protein that binds T3-response elements.";
RL DNA Cell Biol. 15:643-651(1996).
CC -1- FUNCTION: ACTS AS A TRANSCRIPTIONAL REPRESSOR. BINDS TO E-BOX
CC SEQUENCES IN THE IMMUNOGLOBULIN HEAVY CHAIN ENHANCER AS WELL AS IN
CC THE REGULATORY REGIONS OF MANY OTHER TISSUE-SPECIFIC GENES (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=zfhep-1;
CC IsoId=Q62947-1; Sequence=Displayed;
CC Name=2; Synonyms=zfhep-2;
CC IsoId=Q62947-2; Sequence=VSP_006881;
CC -1- SIMILARITY: BELONGS TO DELTA-EF1/ZFH-1 FAMILY OF TWO-HANDED ZINC
CC FINGER/HOMODOMAIN PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U51583; AAB17130.1; -;
CC EMBL; U51584; AAB17131.1; -;
CC InterPro: IPR001356; Homeobox.
CC InterPro: IPR007087; Znf_C2H2.
CC Pfam: PF00096; Zf-C2H2; 7.
CC DR ProDom; PD000010; Homeobox; 1.
CC DR ProDom; PD000003; Znf_C2H2; 2.
CC DR SMART; SM00389; HOX; 1.
CC DR SMART; SM00355; Znf_C2H2; 7.
CC DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
CC DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 6.
CC KW Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;
CC KW Homeobox; Repressor; Activator; Metal-binding; Repeat;
CC KW Alternative splicing.
CC FT ZN_FING 150 173 C2H2-TYPE.
CC FT ZN_FING 180 202 C2H2-TYPE.
CC FT ZN_FING 220 242 C2H2-TYPE.
CC FT ZN_FING 248 272 C2H2-TYPE (ATYPICAL).
CC FT ZN_FING 559 618 HOMEBOX-LIKE.
CC FT ZN_FING 881 903 C2H2-TYPE.
CC FT ZN_FING 909 931 C2H2-TYPE.
CC FT ZN_FING 937 958 C2H2-TYPE (ATYPICAL).
CC FT DOMAIN 968 1109 GLU-RICH (ACIDIC).
CC FT DOMAIN 1 198 Missing (in isoform 2).
CC FT VARSPLIC 1 198 /FtId=VSP_006881.
CC SQ SEQUENCE 1109 AA; 121626 MW; 12162626 BEPE251C8795D6A6 CRC64;
Query Match 31.1%; Score 1550.5; DB 1; Length 1109;
Best Local Similarity 39.2%; Pred. No. 2e-66;
Matches 387; Conservative 150; Mismatches 294; Indels 157; Gaps 35;
QY 2 LTOGAGRRKFKCECGAKFYKHLKHLRIHSGEKYECPCNCKKRSFGSSVSHISSK 61
DB 211 VTSGGKRRKCKCEGAFYKHLKHLRIHSGEKYECPCNCKKRSFGSSVSHISSK 270
QY 62 KCIGLISVNGRMNNIKTG--SSPNVSASSPTNSAITQILRNKLENGKPLMSBOTGLDKI 119

DB 271 KGISLMPVNGRRSGGLKTSQCSSP-SLSTSPSPRPPIRQIKTEN-KPL--DEPLSVNQI 326
QY KTEPLDENYKYLMTHTGSGSSPFMNG---GLG---ATSPLGHPASQSPMQLGVGM 172
DB 327 KTEPDVY-EFKETVVASGNCSTPLONGVFGGGLQATSTSPQGVAVLPT----VGL 381
QY 173 EAPLGFPPYMNLSLEVQKVLQI-VDNTVSRQKMDKTEDISKLGYNHMDPCSQPEEG 231
DB 382 VSP-----STNLSDIQVLKVAVDGNIIRQVLENNQASL-----ASKEGPA 423
QY 232 VTSPPN-----IPVGLPVVSHNGATYSITDYLEKNEKACLOSTTTSRQISNI 283
DB 424 VASSIQGGSHSVIASISLPLVDQGTITIIINYSLEGGQLQVVPQNLKKNPAPPKSC 483
QY 284 KKKKRLTLDLTVD-DKMIENHSISPPSCQCKSPFPI-PLHOHERYLKMMBEIKA 341
DB 484 KSEKSPEDLTVRSKQKSPFGAADS--TCLLC-EDCPEDLALFELKHY----- 530
QY 342 VLQPHENIVPNKAGVFDNKKALLSSVLSSEKGLTSPIN-PYQDHSVLYKAYYAMNMPNS 400
DB 531 --DEHPAPPPAPAPATEKPESSASS--AGNDLSPSQPLKNLILKAYYALNAQPS 586
QY 401 DELKISTAVGLPQEFYKWEFQKRYQYNSRSLEFTS-----KPLAP 446
DB 587 EETIKIASVNLPLDVKKKPEKMQAGQIPGQ---SLEPPSGPSGNIPTATEEOPPV 643
QY 447 NSNPPTKDSLRSPPKPMDSITSPSIAELHNSVTSQDPLRLTSSHTNTKAVDKLDH 506
DB 644 DGNEDQEDSTRCQSPKMTSSPVL-PVGSAINSSC----- 679
QY 507 SSNTSPPLNLSSTSKNSHSSSYTPNSPSSSELAEPDLSPKQMBPKGIATKNT 566
DB 680 --TSPPPLNLSAANPQYS---CVSEGTQEPQVEPLDLSLPKQGS----- 723
QY 567 KATSLNLDHNSVSSSENS---DEPLNLTIKKFSPNSNLDKSNNPVYGANPFSK 621
DB 724 -----LLERSITVSYYQNSVSVQEPPLNLSCKKPEQKDSQVTD--SEPVNVVPPSAN 776
QY 622 PLTTPLPQSAFPAPATFMPVPVQTSIPGLRPYGLDMSFLPHAAVYYPGATPADMOOR 681
DB 777 PIVTALPTVTAQLPTTIVAIADQNSVCLPALANKOTIILPQVAYTY---SATVSPAME 833
QY 682 R-KYQKQKQFQDLDGAQDYMGLDWTDSCLSRKKIKKTESGMVACDLCDKTFQK 739
DB 834 PPKVVIQPNQNDERQDTSSEGS-VEDQNDSDCPRPKKTKRAENGVYACDLCDKTFQK 892
QY 740 SSSLRHAKTEHNGKPRPHQCIQKAPKKGHLIHSRLHSGEKPYQDCGGRFSHSGY 799
DB 893 SSSLRHAKTEHNGKPRHPEGCIQKAPKKGHLIHSRLHSGEKPYQDCGGRFSHSGY 952
QY 800 SOHNNRYSYCKREAREAREAREAREK-----HLGPTLLMNRAYLQITPQGSDE 854
DB 953 SOHNNRYSYCKRGADRAMEQE--DTGPALPVLPTLVGARA-----SPSQADSD 1004
QY 855 ERESMPR--DGESEKEKEGEGYQGLRRDQDEEEEBEESSEKNSMDTPETIRDEE 912
DB 1005 ERSLSLRREDEDESEKKEEBEDKEMBELQEDKCNQEBEEBEEBEEBEEBE--EERA 1062
QY 913 ETGDHSMDDSSDEGKMETYSDEHEDNME 940
DB 1063 EEAHEHAALAKTGAVEEBAAOAGSFQ 1090
RESULT 8
ZFH1_DROME
ID ZFH1_DROME STANDARD; PRT; 1060 AA.
AC P28166;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 1 (zinc finger homeodomain protein 1).
ZFH-1.

OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydriidae; Drosophilidae; Drosophila.
 NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92001539; PubMed=1680376;
 RA Fortini M.E., Lai Z., Rubin G.M.;
 RT "The Drosophila zfh-1 and zfh-2 genes encode novel proteins
 containing both zinc-finger and homeobox motifs.";
 RL Mech. Dev. 34:113-122(1991).
 CC -1- FUNCTION: INVOLVED IN THE DEVELOPMENT OF THE EMBRYONIC CENTRAL
 CC NERVOUS SYSTEM, EMBRYONIC MESODERM AND ADULT MUSCULATURE.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- TISSUE SPECIFICITY: MESODERM AND MESODERMALLY-DERIVED STRUCTURES
 CC IN THE EMBRYO INCLUDING THE DORSAL VESSEL, SUPPORT CELLS OF THE
 CC GONADS, AND SEGMENT-SPECIFIC ARRAYS OF ADULT MUSCLE PRECURSOR.
 CC ALSO IDENTIFIED IN MOTOR NEURONS OF DEVELOPING CNS.
 CC -1- SIMILARITY: BELONGS TO DELTA-EF1/ZFH-1 FAMILY OF TWO-HANDED ZINC
 CC FINGER/HOMEOBOX MAIN PROTEINS.
 CC -1- SIMILARITY: Contains 9 C2H2-type zinc fingers.
 CC -1- SIMILARITY: Contains 1 homeobox domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sb.ch/announce/>
 CC or send an email to license@isb-sb.ch).
 CC -----
 DR EMBL: M63449; AAA2050.1; -.
 DR PIR: S33641; S33641.
 DR HSSP: P08153; 1ZED.
 DR TRANSFAC: T00919; -.
 DR FLYBase: FBgn0004606; zfh1.
 DR GO: GO:0007498; P:mesoderm development; IEP.
 DR GO: GO:0007399; P:neurogenesis; IEP.
 DR InterPro: IPR001356; Homeobox.
 DR InterPro: IPR007087; Znf_C2H2.
 DR Pfam: PF00046; homeobox; 1.
 DR Pfam: PF00096; zF-C2H2; 9.
 DR ProDom: PD000010; Homeobox; 1.
 DR ProDom: PD000003; Znf_C2H2; 2.
 DR SMART: SM00389; HOX; 1.
 DR SMART: SM00355; Znf_C2H2; 9.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS00021; HOMEBOX_2; 1.
 DR PROSITE: PS00028; ZINC FINGER C2H2_1; 6.
 DR PROSITE: PS00157; ZINC FINGER C2H2_2; 9.
 KW Zinc-finger; Metal-binding; DNA-binding; Homeobox; Nuclear protein;
 KW Repeat.
 FT ZN_FING 74 97 C2H2-TYPE 1.
 FT DOMAIN 222 252 GLN-RICH (OPA-REPEAT).
 FT ZN_FING 225 317 C2H2-TYPE 2.
 FT ZN_FING 330 352 C2H2-TYPE 3.
 FT ZN_FING 351 383 C2H2-TYPE 4.
 FT ZN_FING 389 413 C2H2-TYPE 5.
 FT ZN_FING 413 457 C2H2-TYPE 6.
 FT ZN_FING 634 657 C2H2-TYPE 6.
 FT DNA_BIND 705 764 HOMEBOX.
 FT ZN_FING 973 995 C2H2-TYPE 7.
 FT ZN_FING 1001 1023 C2H2-TYPE 8.
 FT ZN_FING 1029 1050 C2H2-TYPE 9.
 FT ZN_FING 1060 11413 MW; 3D96A150B03F0567 CRC64;
 SQ SEQUENCE

Query Match 16.4%; Score 819.5; DB 1; Length 1060;
 Best Local Similarity 29.2%; Pred. No. 8, 2e-32;
 Matches 256; Conservative 100; Mismatches 268; Indels 253; Gaps 33;

QY 9 RKFKEGCKAKFYKHLKEHLRIISGEKPYECPCCKKRFSSGSSYSSHSKCCIGLIS 68
 DB 359 RKFKEGCKAKFYKHLKEHLRIISGEKPYECPCCKKRFSSGSSYSSHSKCCIGLIS 415

QY 69 VNGHRRNN-----IKTGSSPNSSSSPTNSAITOLRNKLN-----GKPLSNSEOTGLIKI 119
 DB 416 MGLKLNRRALKRLKLEKSPGSSASSARRSPDHKGKULPEQPSLPLGLPHMS----- 467
 QY 120 KTEPLDENDYVLMATGFCSSGSPPMNGGLGATSP-LGVHPSAOSPMOHLGVGEAPILG 178
 DB 468 -----YFASDAQVOGSGSAAPAPFPFPHDNY-----NMAALLA 499
 QY 179 FP-----TNNSNLS-----EVQXVLQIVDYNVSSQXMDCKTEDISKLGYNMKPCSGPE 228
 DB 500 FPHNFMALAAAGLDPRVHPYSTQRLQL--SAAGQQQREEREEOQKQOH--DEEETPD 554
 QY 229 EGVATSPNIPVPGPVVSHNGATSKIIDYTLKYNKAKACQSLTTSRRQISNIKEKL 288
 DB 555 EPKLVMDIEEETEMAPTPATEAATPIKREBERA-----SDPEYRSSQAIKEQ 609
 QY 289 RTLIDLVTDKMIENHS--ISTPPSCQCKESPFGPIPLHQRHYLCRMNEETKAVLQ 345
 DB 610 EPLVAERQTPVEHHAPEHSAADLRCSRCSPFNFHPTLVQHEKVLGCL--IKELEQ 666
 QY 346 HENIVPKAGVFN-----DNKALLSSVLSEKQ-----LTSPINPYKDHNSVYKA 390
 DB 667 H-FOQQAITSFALASASEDEDEBEMDVEEPEQSEKRVARTAIN--EQQQQQLKQ 722
 QY 391 YYAMNMEPNDELKISIAVGLPOEFYKEMPE-----QRXYOYSNRS----- 434
 DB 723 HYSLNARSRREFPMIARLQLDPRVYQVQNNRNRKRNQSTQNNQAAGAAPMPID 782
 QY 435 -PSLERTSKPLAPNSNPTTKOSLLPRSPVKPMDSITSPSIAELHNSVTSQDPLRLTKSS 493
 DB 783 QASLTRDQPL-----DLVARDPLTPKSESSPYIA-----PP-----SG 818
 QY 494 HFTNPKAVD-KLDSRSNTSPPLSLSTSSSKNSHSSSTTPMSFSSEELQAP----- 544
 DB 819 EALNPEALNISRRKSTSSMSPPASISPSAALYFGAAPSPSPNSQLDSTPREGQAPEG 878
 QY 545 -----LDLSLPKQ-----MRBPKGIATGNKTKATSYNL--DHNSSSSSENDEPLNLF 594
 DB 879 LPPYMLPMSLMELAFKRR-PGDFPAS-NHALLMSIKLPDTRGTSLSFGSGE----- 928
 QY 595 KKEFSNSNNLDNKSNNPFGNPPSAKPLVTPLPQSAFPPTAMPVQVTSIPGLRPYGP 654
 DB 929 KRSWRDD-----SRISHEDFAGVLMPP----- 953
 QY 655 LDQMSFLPMAYTYPTGAATPADMQRRKYORKGFGQDLDGADYDMSGDMDTDSNC 714
 DB 954 -----KPRGKYETHGHAD-----PDLP----- 972
 QY 715 LSRKKIKTESGMYACDLCDKTFQKSSSLRHKYTEHTGKPRHOCICCKAFKHHLIEH 774
 DB 973 -----YVQDCQDAFAKQSSLAHKKIEHSQRRYQCLIECFKAKKHHLIEH 1019
 QY 775 SRLHSGEKPYQCDKCGKRFSSGSSYQHMNRYSYCK 811
 DB 1020 KRLHSGEKPYQCSKCLKRFSSGSSYQHMNRYSYCK 1056

RESULT 9
 ID 2236 HUMAN STANDARD; PRT; 1845 AA.
 AC 09UL36; 09UL37;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein 236.
 GN ZNF236.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

CC TISSUE=Kidney; PubMed=10458916;
 CC MEDLINE=93389731; Mason R.M.;
 CC Holmes D.I.; Wahab N.A.;
 CC "Cloning and characterization of ZNF236, a glucose-regulated kruppel-
 CC like zinc-finger gene mapping to human chromosome 18q22-q23.";
 CC Genomics 60:105-109(1999). AS A TRANSCRIPTION FACTOR.
 CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=B;
 CC IsoId=Q9UL36-1; Sequence=displayed;
 CC Name=A;
 CC IsoId=Q9UL36-2; Sequence=VSP_006907, VSP_006908;
 CC -1- TISSUE SPECIFICITY: UBIQUITOUS. EXPRESSION LEVELS ARE HIGHEST IN
 CC SKELETAL MUSCLE AND BRAIN, INTERMEDIATE IN HEART, PANCREAS, AND
 CC PLACENTA, AND LOWEST IN KIDNEY, LIVER, AND LUNG.
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF085244; AAD55329.1; -;
 CC EMBL: AF085243; AAD55328.1; -;
 CC HSSP: P0746; IARD.
 CC GeneW: HGNC:13028; ZNF236.
 CC MIM: 604760; -;
 CC DR GO: GO:0005634; C:nucleus; NAS.
 CC DR GO: GO:0003700; P:transcription factor activity; NAS.
 CC DR GO: GO:0006355; P:regulation of transcription; DNA-dependent; NAS.
 CC DR InterPro: IPR007087; Znf_C2H2.
 CC DR Pfam: PF00096; Zf_C2H2; 30.
 CC DR ProDom: PD000003; Znf_C2H2; 4.
 CC DR SMART: SM00355; Znf_C2H2; 30.
 CC DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 30.
 CC DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 30.
 CC KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 CC Nuclear protein; Repeat; Alternative splicing.
 CC FT ZN_FING 37 59
 CC FT ZN_FING 66 88 C2H2-TYPE.
 CC FT ZN_FING 93 115 C2H2-TYPE.
 CC FT ZN_FING 121 143 C2H2-TYPE.
 CC FT ZN_FING 153 175 C2H2-TYPE.
 CC FT ZN_FING 197 219 C2H2-TYPE.
 CC FT ZN_FING 225 247 C2H2-TYPE.
 CC FT ZN_FING 253 276 C2H2-TYPE.
 CC FT ZN_FING 285 308 C2H2-TYPE.
 CC FT ZN_FING 482 504 C2H2-TYPE.
 CC FT ZN_FING 510 532 C2H2-TYPE.
 CC FT ZN_FING 538 560 C2H2-TYPE.
 CC FT ZN_FING 566 588 C2H2-TYPE.
 CC FT ZN_FING 657 679 C2H2-TYPE.
 CC FT ZN_FING 685 707 C2H2-TYPE.
 CC FT ZN_FING 713 735 C2H2-TYPE.
 CC FT ZN_FING 741 763 C2H2-TYPE.
 CC FT ZN_FING 967 989 C2H2-TYPE.
 CC FT ZN_FING 995 1017 C2H2-TYPE.
 CC FT ZN_FING 1023 1045 C2H2-TYPE.
 CC FT ZN_FING 1051 1073 C2H2-TYPE.
 CC FT ZN_FING 1166 1189 C2H2-TYPE.
 CC FT ZN_FING 1167 1189 C2H2-TYPE.
 CC FT ZN_FING 1195 1217 C2H2-TYPE.
 CC FT ZN_FING 1223 1245 C2H2-TYPE.
 CC FT ZN_FING 1251 1273 C2H2-TYPE.
 CC FT ZN_FING 1657 1680 C2H2-TYPE.
 CC FT ZN_FING 1686 1708 C2H2-TYPE.
 CC FT ZN_FING 1722 1744 C2H2-TYPE.

FT ZN_FING 1750 1772 C2H2-TYPE.
 FT ZN_FING 1778 1801 C2H2-TYPE.
 FT VARSPLIC 1530 1558 ELNATTSGLPSTPTTPSPASISTONLWMS -> GSRVQHS
 FT VARSPLIC 1559 1845 VGOECGSVEALYLENSSDKT (in isoform A).
 FT VARSPLIC 1559 1845 /FTID=VSP_006907.
 FT VARSPLIC 1559 1845 Missing (in isoform A).
 FT VARSPLIC 1559 1845 /FTID=VSP_006908.
 SQ SEQUENCE 1845 AA; 203659 MW; 2879EA91D0C6DD3D8 CRC64;
 Query Match 8.2%; Score 407; DB 1; Length 1845;
 Best local similarity 20.6%; Pred. No. 5.8e-12;
 Matches 220; Conservative 133; Mismatches 385; Indels 370; Gaps 49;
 QY 7 GNRKFKTECGAKFKYKHLKEHLRHSGEKPECPNCKRFSHSGSYSHSKKICGL 66
 DB 221 GERPFCSGCGAFNKGALQTHMIKHTEKHACFCFPAASOKGNLQSHVQ----- 273
 QY 67 ISVNGMRNNTITGSSPNSVSSPTNSALTQIR--NLENGKPLMSQTLGKI----- 119
 DB 274 -----RVHSEVKGPTYNCTECSVFKSLGSLNTHISKMHGQPONSTSTETAHLVLTAT 328
 QY 120 --KTEPLDNDYKVLMTATGFGSG--SPMNGGLATSPGVHPASQSPM--QHLG--VG 171
 DB 329 LQGTLPLOQTEAATSAASQPSQAVSDVIQQLLESEAPVE--SQGSPQGOQLSITVG 387
 QY 172 MEAPILGFETMNSNLSEVQKVLQIVDNTVSRQKMDCKTEDISKLGYHMKDPC---SQP 227
 DB 388 INDILQALENSGLSIPAA-----HPNDSCHAKTSAP 422
 QY 228 EEOGVTSPNIPVGVLPVSHNGATSIIDYLEKNVEAKCYQLTSDSRRIQINIKERK 287
 DB 423 HAO--NPDVSVS-----NE-----QTPEDAEQKEQESPERK 453
 QY 288 LRTLIDLVDDKMIENHSISTPFS-----COFKESFPGP-----IPLHCHER 330
 DB 454 LDK-----KEKMIKKKSFPFGSIRENGVNMHCYCAKFRKPSDLVRIHITHRK 508
 QY 331 YL-----CKNBEIKAVLQPHENIVPNKAGFVQNKALLISVLESG-----LTSPIIN 379
 DB 509 PFKPCQCFPAFAVAKSTLTAA---IKTHGIIKAFKQCYCKKS--FSTGSLKVIHLRTGVVR 564
 QY 380 PYQDHNSVLKAYANMMEPNSDELKISTAVGLPQEFVEMEQRRVYQYNSRSLSLR 439
 DB 565 PFC-----PHCDKFKPTS---GHRKTHLASHKHTELRMRQIRKPAKVR 607
 QY 440 TSKPLAPNSNPTTKOSLPRSPVFKPMDSITSPSIAE-----LHNSVTSQDPL 487
 DB 608 VKTINVP-----VPDIPLOEPILITDGLQIPRKQPFQSYFNNNVYNEADRY 657
 QY 488 R-----LTKSHFTNIKAVDKLHRSNT--PSPILNIS-----STSSKNSHSSSYTP 532
 DB 658 KCFYCHRAVKKSCHLK-----QHIRSHTEKPKKSCQCRGFVSAGVLKAIHITHVG 709
 QY 533 -NSFS-----SEEL-----QAEPLDLSLPQMBEPKIINT 562
 DB 710 LKSFKCLIONGAFTTGSLRRHMGIHNDLRPYMCPYCOQTFPTSLNCKKMHKTHRELAQ 769
 QY 563 KNTTKATSLNLDHNSVSSSENSDEPLNTFTIKKEPSSNNLNDKNSNPVFCMNP----- 617
 DB 770 QLOQHQAASIDSTYDQSQMQASTQMOVEIESDELPTAEVVAANPEAMLDLEPQHYVG 829
 QY 618 -----FSAKPL-----YTPPL-----PQASAPPA-----T 637
 DB 830 TEEAGLGOQLADQPLEADBDGFVAPQDPLRGHVDCFEQSPAQQSPEPAGLPQGFVTDT 889
 QY 638 F-----MPVQ-----TSI--PGLRPYGLDQMSFLPHMAVYTPGTATRA 676
 DB 890 YHQQPPFPVQQLQDSSTLESQALSTSFHQSLQAQSSDGANNVTRL-----IQ 939
 QY 677 DMQQRKRYQKQGFQDLDLDAQDVYSGLDMDTSDSCSRKKIKKTESGMYACDLCDKT 736
 DB 940 ESSQEBELDLQAQSSQ-----FLEDNEDQ-----SRNS-----YRCDICNKG 975

Query Match	Best Local Similarity	Matches	Score	DB	Length
7 GNRKFKCECGCAFYKXKHLIKEMHLRIHSGEKDYECPCNCKKRFSSSSSYSHSSKXCIGL	21.5%	214	7.6%	340	1191
290 GEEPYKCECGCAFHSSTLAKHRIHGEKPYKCECGCAFSSSALAH	Conservative 115	Mismatches 365	Pred. No. 7.1e-11	Indels 300	Gaps 39
67 ISVNGRMNRNITKYGSSPN	SVSSSPNTSAITOLRNK	101			
341 KRITTGKPYKCKECCGAFNSSTLANHKITHEEKPYKCKECDKTFKRLST	392				
102 LKNGKPLNSBOTGLIKITTEPLDNDYKVLMTATG	151				
393 LTKHKLIHAGEKLYKCECGKA	443				
152 TSPKGYH	206				
444 NWSSTLTTRKRPHTREKPFKCECK	483				
207 CTETDTSKLGYHMDPCSOPEEOGVS	266				

```

Db 484 HNGE-----KPYKCEB-CGKAFROSST-----LTHGKIIHTEKPKYKE 521
QY 267 ACLQSLTDSRRIISNIKKKELATLDVTDGMIENHSITPSPCOFCSEPGPIPLH 326
Db 522 ECGKAF-----KQSLTLNKHKI-----IHSKEKKYKCECKAKAKOSTLT 562
QY 327 QHE-----RYLCQMEIKAV-----LOPHENIVPKAGVFD--NKALLLSVL--SEKG 373
Db 563 THKIIHAGKKLYKCECGKAFNHSLSLTHKI IHTGEKSKYKCECGKAFIWSSTLRRHKR 622
QY 374 LTPSPINPYK-----DHMSVLKAYAMMEBNSBELKISIAVLPGQFYKWEQOK 425
Db 623 IHTGEKPKYKCECGKAFSHSSALAKH-----KRIHTEK 656
QY 426 VVO-----YSNRSRPSLERTS-----KPLAPNSNPTTKOSLLPRSPVKPMDISPSIA 474
Db 657 PYKCECGKAFSSSTLANHKKIHTBEKPKYKCECKT-----FKRLSTLTHKII 707
QY 475 ELHNSVTSQDP--PLRLTKSSHPTNIRKAVDLHRSRNTSPSLNLSSTSSKNSSSYTPN 533
Db 708 HNGEKLKYKCECGKAFNRSNLT-----IHKFIHTGE--KPYKCECGKAFNHSLSLTH 760
QY 534 S--FSSEELQAEPLDLSLPKQMB--PKGIATKNTKATISINDHNSVSSSSSENSDEPLN 590
Db 761 KRITREK-----PFKCEKCEGKAFIWSSTLTHKRIH-----TGEKPKYK 799
QY 591 LPTIKKEFSNSNLDKSNPNVFGMNPFAKPLTYPLPQSPAPPATFMPVOTSIPLGR 650
Db 800 CEKCGKAFRSSTLT--TKKTIHTGEKPKYKCEC-----GKAKHSSALAKHKKI IHAKEK 852
QY 651 PY-----PGLDQMSFLPHNAVYTPYTGAAFPAD-----MOQRKRYOKQOFQ 692
Db 853 LYKCECGKAFNQSNTLTHKIIHTKEKPKSECDKAFIWSSTLTHKRIHTEKPKYK 912
QY 693 DLDGQODYWSGL-----DMTDSLSLKKIKYKTESGMYADLCD 734
Db 913 BECGKAFSOPSHLTTHKRMHTGKPKYKCECGKAFSOSSTLTHKIIHTGEKPKYKCECG 972
QY 735 KTKQKSSSLRHYEHYHTGKPKPHOCQICKAFKPKHHLIEHSRLHSGEKYQCDKCGKRF 794
Db 973 KAFKSSSTLTHKIIHTGEKPKYKCECGKAFSOSSTLTHKRMHTGKPKYKCECGKAFN 1032
QY 795 HSGSYSOHMHRYSYCKREAEEREAERAREKGLAPTELLNRAVYLOSITPOGYSDE 854
Db 1033 RSEKLTTHKIIHTGEKPKYKCE-----CGKAFISSTLNGHKRIH 1072
QY 855 ERESMRDGESEK-----EHEK--EGEGY 877
Db 1073 TREKPKYKCECGKAFSOSSTLTHKRLHTGEKPY 1106

RESULT 11
EVIL_HUMAN STANDARD; PRT; 1051 AA.
ID EVIL_HUMAN
AC 003112; Q16122; Q99917;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ecotropic virus integration 1 site protein.
GN EVIL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90326419; PubMed=2115646;
RA Morihata K., Parganas B., Douglass E.C., Ihle J.N.;
RT "Unique expression of the human Evi-1 gene in an endometrial
RT carcinoma cell line: sequence of cDNAs and structure of alternatively
RT spliced transcripts."
RL Oncogene 5:963-971(1990).

```

```

RN [2]
RP SEQUENCE OF 970-1012 FROM N.A.
RX MEDLINE=96292241; PubMed=8700545;
RA Ogawa S., Kurokawa M., Tanaka T., Mitani K., Inazawa J.,
RA Hangishi A., Tanaka K., Matsuo Y., Minowada J., Tsubota T.,
RA Yasaki Y., Hirai H.;
RT "Structurally altered Evi-1 protein generated in the 3q21q26
RT syndrome."
RL Oncogene 13:183-191(1996).
RN [3]
RP VARIANT AML1-EVI1 FUSION IN CHRONIC MYELOCYTIC LEUKEMIA.
RX MEDLINE=94147997; PubMed=8313895;
RA Mitani K., Ogawa S., Tanaka T., Miyoshi H., Kurokawa M., Mano H.,
RA Yasaki Y., Ohki M., Hirai H.;
RT "Generation of the AML1-EVI-1 fusion gene in the t(3;21)(q26;q22)
RT causes blastic crisis in chronic myelocytic leukemia."
RL EMO J. 13:504-510(1994).
CC -1- SUBUNIT: MAY INTERACT WITH CTBP1 (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=Long;
CC IsoId=Q03112-1; Sequence=Displayed;
CC Name=2; Synonyms=Short;
CC IsoId=Q03112-2; Sequence=Not described;
CC -1- DISEASE: INVOLVED IN CHRONIC MYELOGENOUS LEUKEMIA (CML) BY A
CC CHROMOSOMAL TRANSLOCATION T(3;21)(Q26;Q22) THAT INVOLVES EVI1 AND
CC PROTEIN AML1.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X54989; CAA8735.1; -
DR EMBL; S82592; AAB37456.1; -
DR EMBL; S69002; AAB32907.1; ALT_SEQ.
DR PIR; A60191; A60191.
DR PIR; S41705; S41705.
DR HSSP; P08047; ISP2.
DR GeneW; HGNC:3498; EVI1.
DR MIM; 165215; -
DR GO; GO:0005634; C:nucleus; NAS.
DR InterPro; IPR007087; Znf.C2H2.
DR Pfam; PF00086; Zf-C2H2; 10.
DR ProDom; PD000003; Znf.C2H2; 4.
DR SMART; SM00355; Znf.C2H2; 10.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 10.
KW Zinc-finger; Metal-binding; DNA-binding; Developmental protein;
KW Proto-oncogene; Chromosomal translocation; Nuclear protein; Repeat;
KW Alternative splicing.
FT ZN_FING 21
FT ZN_FING 75
FT ZN_FING 103
FT ZN_FING 131
FT ZN_FING 160
FT ZN_FING 188
FT ZN_FING 217
FT ZN_FING 239
FT ZN_FING 733
FT ZN_FING 761
FT ZN_FING 790
FT ZN_FING 812
FT DOMAIN 421
FT DOMAIN 434
FT SITE 886
FT SITE 937
FT SITE 953
FT SITE 984
FT SITE 584
FT SITE 588
FT CONFLICT 303
FT CONFLICT 303
FT CONFLICT 672
FT CONFLICT 672
FT CONFLICT 796
FT CONFLICT 796
FT CONFLICT 875
FT CONFLICT 875

```

FT CONFLICT 861 861 P -> T (IN REF. 3).
 FT CONFLICT 906 906 Y -> N (IN REF. 3).
 FT CONFLICT 1008 1012 AYAMM -> VOIFP (IN REF. 2).
 SQ SEQUENCE 1051 AA; 118335 MW; 3270955E25D99D51 CRC64;

Query Match 7.1%; Score 355.5; DB 1; Length 1051;
 Best Local Similarity 19.4%; Pred. No. 7.9e-10;

Matches 199; Conservative 140; Mismatches 395; Indels 293; Gaps 36;

9 RRFKTEGCGAKFYGHLEHRLHSGEKYECPCNKKRFSHSGSYSHISKCCIGLIS 68
 101 REYKDCQPKAFNWSNLIRHOMSHDSKHYECACAKFTDPSNLORHINSQHV----- 155
 69 VNGRMNNIKTGSSPNVSSSPTSNSAITQLRNKLNGKPL-----SMSEQTGLIKIT 121
 156 -----GARAHACECGKFTATSSGLQKHIIHSVXPFICEVCHKSYTQSNLCRHR 208
 122 EPLD-----FNDYKVLMAITGF--SGSPFMNGCIGATSPUGVHPSAOSPQOH 167
 209 MHADRTQIKCKDCQCMFSTTSLNKHRRFCGKHHFAAGGF----- 250
 168 LGVGEAPLPGPTTNSNLSEVQYVLQIVDNTVSRQKDCKTEDISKLGKYMKDPGSGP 227
 251 FGQGISLP--GTPAMDK-----TSVMNNSHANPGIADYTFG-----ANR 286
 228 EEOGVTSNPINP-----PVGH-----PVVSHNGATKSIIDYTLKVNKAQCLQSILT 273
 287 HPAGLTPTAPGFSVPGLPFGSLYHRLPLPASPVKGL--STEQNNSSQS----- 338
 274 TDSRQISNIKKKRLTLIDVTDNMIENHSITPFCQCFKESFPGPPIPLHGHRYLC 333
 339 -----PLMTHPQILP 348
 334 KMMNEIKAVLOPHENIVPNKAGVAFVFNKALLSVLSEKGLTPIPNPKDMSVLKAYYA 393
 349 ATQDLTKA--LSKHPSVGDNK-----EYELOPERSSEERPEK 384
 394 NMMEFNSDELKISTAVGLPOEFVKEMPEQRKYQVSNRSRPSLERTSKPLAPNSNPTTK 453
 385 ISDSSESDLDVDSVTPSGSDELT-----SGSDLESIDESDEKKEKENGKPK 432
 454 DSLPRSPVKKPMD-----STSPSIAE--LHNSVTSQDPLRLTKSSHFTNIKAVD 502
 433 DKVSPLOMLASINNKKESYNNHSIFPSPLEEQTAVGAVNDISIKALASIAEKFGSGTGVG 492
 503 KLDHRSMTPTPLNLSSTSSKSHSSSYTPNFSSEELQAPBLD-----SLPQOM 553
 493 LQDKKVGALPYR--SMFPLPFPFAPFSQSWYP--FPDRDLRSLLPKMEPOSPEGVKLQKGS 549
 554 REPKGIIATKNTKATKATINLDHNSVSSSENSEDEPLNLTFFICEKFSNNSNLDNKNPNPVF 613
 550 SESPPDLTTKRDEKPLTPVPSPKPVTPATSDQDLDSMGRSRAASGKLTPEPKNHVF 609
 614 G-----MNPFSAKPLYTLPPLPQSAAPPATPMPV-----QTSIP-----GLRPY 652
 610 GSKKGSNVSERAPASDGL-----QHARPTPEFMDPIYREKRLTDLPLEALKEXYLPBS 663
 653 PGLDMSLPLPHAYTYTTCGATFADMQRKRYQKQKQGGDLDDAODYMSGLDMTPD-- 710
 664 PG-----FLFHFQFLPQORTMSAIEENNA--EKLESFSA--LKPASSELQSVPMFNR 715
 711 -SDSCLSRRKIKTTSGMYACDLCKTFQKSSSLRHKHEHNGKRRPHOCQKKAFFKHG 769
 716 APPNALPENLARKGE--RYTCRYCGKIFPRSANLRLHRLTHGEOPYCKYCDRSPFSIS 774
 770 HLIEHSR--LHSGEKYQCDKCGKRFSGSHSGSYQHNNHRYSYCKRAEEREAERAREK 828
 775 NLOHRAVRIHNKEKPKCHLCYRCFGQGNLDRHL-----KXHEHG 815
 829 HLGPELLMNRAYLOSITPQGSYDSEERSMPPDSESEKHEKE--GREGYKGLRRDG 885
 816 NMSTG-----ATSSPHSELESTGAILLDKEDAYFTIRNFTGNSNHSQSPRNV 864

QY 866 DE-----EEEEEESENKSM---DTDEPTRDEEETDHSMDSSDEGKMETYSOH 934
 DB 865 EERNNGSHFKEKXALVPQNSDILLDEVEDEVLDEE---DEBYDITGKTGKEPVTNTL 921
 QY 935 EEDNMD 941
 DB 922 HEGNPD 928

RESULT 12

XFIN XENLA STANDARD; PRT; 1350 AA.
 ID P08045;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein Xfin.
 GN XFIN.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 OC Xenopodidae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=88082679; PubMed=2826129;
 RA Ruiz i Altaba A., Perry-O'Keefe H., Melton D.A.;
 RT "Xfin: an embryonic gene encoding a multifingered protein in
 RT Xenopus";
 RL EMBD J. 6:3065-3070(1987).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=94021366; PubMed=7692399;
 RA Andreatzoli M., de Lucchini S., Costa M., Barsacchi G.;
 RT "RNA binding properties and evolutionary conservation of the Xenopus
 RT multifinger protein Xfin.";
 RL Nucleic Acids Res. 21:4218-4225(1993).
 RN [3]
 RP STRUCTURE BY NMR OF FINGER 31.
 RX MEDLINE=89346749; PubMed=2503871;
 RA Lee M.S., Gipeert G.P., Soman K.V., Case D.A., Wright P.B.;
 RT "Three-dimensional solution structure of a single zinc finger DNA-
 RT binding domain";
 RL Science 245:635-637(1998).
 RN [4]
 RP STRUCTURE BY NMR OF A FINGER.
 RX MEDLINE=89378224; PubMed=2506074;
 RA Lee M.S., Cavanagh J., Wright P.B.;
 RT "Complete assignment of the 1H NMR spectrum of a synthetic zinc
 RT finger from Xfin. Sequential resonance assignments and secondary
 RT structure";
 RL FEBS Lett. 254:159-164(1989).
 CC -1- FUNCTION: BINDS RNA. COULD FUNCTION IN POST-TRANSLATIONAL
 CC REGULATION PROCESSES.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: OOCYTES AND IN SPECIALIZED CELL TYPES SUCH
 CC AS NEURAL RETINA CONES.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT OOGENESIS AND
 CC EMBRYOGENESIS.
 CC -1- DOMAIN: CONTAINS 37 FINGER MOTIFS IN 6 DOMAINS.
 CC -1- PTM: PHOSPHORYLATED.
 CC -1- SIMILARITY: BELONGS TO THE KRUPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -1- SIMILARITY: Contains 1 KRAB domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).


```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88311086; PubMed=2842066;
RA Morishita K., Parker D.S., Mucenski M.L., Jenkins N.A., Copeland N.G.,
RT "Retroviral activation of a novel gene encoding a zinc finger protein
in IL-3-dependent myeloid leukemia cell lines.",
RL Cell 54:831-840(1988).
CC - SUBUNIT: MAY INTERACT WITH CTBP1 (BY SIMILARITY).
CC - SUBCELLULAR LOCATION: Nuclear.
CC - CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-11 IS THE INITIATOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M21829; AAA40581.1; ALT_INIT.
DR PIR: A31591; A31591.
DR HSSP: P08047; 1SP2.
DR TRANSFAC; T00273; -.
DR MGD; MGI:95457; Ev11.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 10.
DR ProDom; PD000003; Znf_C2H2; 4.
DR SMART; SM00355; Znf_C2H2; 10.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 10.
KW Zinc-finger; Metal-binding; DNA-binding; Developmental protein;
KW Repeat; Proto-oncogene; Chromosomal translocation.
FT ZN_FING 21 44 C2H2-TYPE.
FT ZN_FING 75 97 C2H2-TYPE.
FT ZN_FING 103 125 C2H2-TYPE.
FT ZN_FING 131 154 C2H2-TYPE.
FT ZN_FING 160 182 C2H2-TYPE.
FT ZN_FING 188 210 C2H2-TYPE.
FT ZN_FING 217 239 C2H2-TYPE.
FT ZN_FING 224 246 C2H2-TYPE.
FT ZN_FING 752 775 C2H2-TYPE.
FT ZN_FING 781 803 C2H2-TYPE.
FT DOMAIN 421 434 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT SITE 877 928 ASP/GLU-RICH (ACIDIC).
FT SITE 553 557 CTBP-BINDING MOTIF 1 (BY SIMILARITY).
FT SITE 584 588 CTBP-BINDING MOTIF 2 (BY SIMILARITY).
SQ SEQUENCE 1042 AA; 116847 MW; 8DEDF164F536D2FE CRC64;
Query Match 6.9%; Score 345; DB 1; Length 1042;
Best Local Similarity 19.9%; Pred. No. 2, 5e-09;
Matches 212; Conservative 140; Mismatches 367; Indels 344; Gaps 47;

```

```

QY 240 VG-----LPVSHNGATKSIIDYLEKANE--AKACLOSLTDSRRQISNIKKELKT 290
DB 363 VQDNKPEVLELP-----ERSSEERPLEKISDPSGSSDLDDVSTPGSGLFTTSGSDLES 415
QY 291 LIDVLTDDKMIENHSITPPSCQPKESFPFGIPLHGHRYLCKRNEELKAVLQHEMI- 349
DB 416 --DLESDEK-----CKEN-----GKMFKDVSPLOINIA 442
QY 350 -VPNAGVFDNKKALLSSVLESEKLTSPIN-PYKDHMSVLKAYAMNNEPNSDELTKIS 407
DB 443 SITNKE--HNHNSVPSASVEQSAVSGAVNDSTAIASIAEKY-----GSTGVGLQ 494
QY 408 I---AVGLPOEFYKWE--QRKYQYNSRSPSLERTSKPLAENSPTTDSLLPSP 461
DB 495 DKVGVALPYPSWFLPFPFPAFSQWYF-----PDRDLRSJPL-----KNEQSP 539
QY 462 VKPMDSITSPSLAEHLNSVTSQDPLRLTKSHFTNICAVDKLDHSRNTSPPLNLSST 521
DB 540 -----SEVKKLQGSSESPPDLTTKRD 562
QY 522 SKNHSSTYTPNSFSEELQAPLDSLQKREPEKGIATKNTKATSIINDHNSVSS 581
DB 563 EKPILTSGPSKSGTPTATS-QDQPLDLSGSKR-----ASGTYLTPPKNHVFGEEK 613
QY 582 SENSD-EPNLNITFIKKERSNNLNDKSNNPVFGNPP---SAKPLYPLPQSAFPAPT 637
DB 614 GSNMDTRF-----SDQSLQHARPTPF-WDPIYRVKRLTLPDLAKK----- 657
QY 638 FMPPVQTSIPGLRPYGL---DQNSFLPMATYPTGATPADMOQRKRYORKQFQGL 694
DB 658 -----KYLRPSPGLFHPQMSAENMA-----EKLSFSA-L 688
QY 695 LDGADVYSGLDMD---SDSCSRKKIKTESGMYACDLCDTKFOKSSILRKYHT 751
DB 689 KPEASILLQVPSMSPAPPTLPLNLRKKE-RITCRVCGKLPFSAVLTRLRTH 747
QY 752 GKRPHQCOICKAFKHKHLIEHSR-LHSGEKPYOCDGKGRFSGSYQHMHNRYSYC 810
DB 748 GEQPRCKYCCRSFSSNLQRHVANHKEKPFCHCDRFQGOQTNLDRL----- 800
QY 811 KKEAEERAEAREAKHGLFTLLMRAVYQST-----TPQY-----SDSEER 856
DB 801 -----KKHENMMSGTATSPHSELESAGAILDDEDEVYFEIRNFINSNHG 848
QY 857 ESMRPDGESEK--BEHKEGSEGYGLARRD--GDEEESE--ESESNNKMDTPE--- 906
DB 849 SOSPPNMEERNNGSHFKD-KKALATSONSDLLDDEVEDEVYLLDEDEDNDIPGRKEL 907
QY 907 -TIRDEETGDHSMDSSEDEGKMETKSD-----HEEDNMEDQM 943
DB 908 GYTRIDEELPE---DYREAGALEMSCKASPVRYKEEDYKGL 947

```

```

RESULT 14
2208 HUMAN STANDARD; PRT; 1167 AA.
AC 043345;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 208.
GN ZNF208 OR ZNF91L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=98391764; PubMed=9724325;
RA Eichler E.B., Hoffman S.M., Adamson A.A., Gordon L.A., McCreedy P.,
RA Lamerdin J.E., Mohrenweiser H.W.;
RT "Complex beta-satellite repeat structures and the expansion of the
zinc finger gene cluster in 19p12.";

```

RL Genome Res. 8:791-808(1998).
 CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- TISSUE SPECIFICITY: Ubiquitous.
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.
 CC -1- SIMILARITY: Contains 1 KRAB domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AC003973; AAB94784.1; -.
 DR HSSP: P25490; IUBD.
 DR GeneW: HGNC:12999; ZNF208.
 DR MIM: 606760; -.
 DR GO: GO:0005634; C:nucleus; NAS.
 DR GO: GO:0003677; F:DNA binding activity; NAS.
 DR GO: GO:0008270; F:zinc ion binding activity; NAS.
 DR GO: GO:0006355; F:regulation of transcription, DNA-dependent; NAS.
 DR InterPro: IPR001099; KRAB.
 DR InterPro: IPR007087; Znf_C2H2.
 DR Pfam: PF01352; KRAB; 1. C2H2.
 DR Pfam: PF00096; zf_C2H2; 33.
 DR ProDom: PD000003; Znf_C2H2; 16.
 DR SMART: SM00349; KRAB; 1.
 DR SMART: SM00355; Znf_C2H2; 33.
 DR PROSITE: PS00805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 33.
 DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 34.
 DR Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 KW Nuclear protein; Repeat.
 FT 2N_FING 201 223 KRAB.
 FT 2N_FING 229 251 C2H2-TYPE (DEGENERATE).
 FT 2N_FING 257 279 C2H2-TYPE.
 FT 2N_FING 285 307 C2H2-TYPE.
 FT 2N_FING 313 335 C2H2-TYPE.
 FT 2N_FING 341 363 C2H2-TYPE.
 FT 2N_FING 369 391 C2H2-TYPE.
 FT 2N_FING 397 419 C2H2-TYPE.
 FT 2N_FING 425 447 C2H2-TYPE.
 FT 2N_FING 453 475 C2H2-TYPE.
 FT 2N_FING 481 503 C2H2-TYPE.
 FT 2N_FING 509 531 C2H2-TYPE.
 FT 2N_FING 537 559 C2H2-TYPE.
 FT 2N_FING 565 587 C2H2-TYPE.
 FT 2N_FING 593 615 C2H2-TYPE.
 FT 2N_FING 621 643 C2H2-TYPE.
 FT 2N_FING 649 671 C2H2-TYPE.
 FT 2N_FING 677 699 C2H2-TYPE.
 FT 2N_FING 705 727 C2H2-TYPE.
 FT 2N_FING 733 755 C2H2-TYPE.
 FT 2N_FING 761 783 C2H2-TYPE.
 FT 2N_FING 789 811 C2H2-TYPE.
 FT 2N_FING 817 839 C2H2-TYPE.
 FT 2N_FING 845 867 C2H2-TYPE.
 FT 2N_FING 873 895 C2H2-TYPE.
 FT 2N_FING 901 923 C2H2-TYPE.
 FT 2N_FING 929 951 C2H2-TYPE.
 FT 2N_FING 957 979 C2H2-TYPE.
 FT 2N_FING 985 1007 C2H2-TYPE.
 FT 2N_FING 1013 1035 C2H2-TYPE.
 FT 2N_FING 1041 1063 C2H2-TYPE.
 FT 2N_FING 1069 1091 C2H2-TYPE.
 FT 2N_FING 1097 1119 C2H2-TYPE.
 FT 2N_FING 1125 1147 C2H2-TYPE.
 FT 1167 AA; 134352 MW; E2184DF23B0D35E9 CRC64;
 SEQUENCE

Query Match 6.8%; Score 341.5; DB 1; Length 1167;
 Best Local Similarity 21.2%; Pred. No. 4.2e-09;
 Matches 190; Conservative 121; Mismatches 313; Indels 271; Gaps 40;
 QY 7 GNRKPCCTCEGKAFKKYKHLKLEHLIHSGEKYECPNCKRRFSHSGSVSHSKKICIL 66
 DB 421 GKKPYKCEGKAFKFWSSNLMKHKIHTGETPYKCECGKFSWSTLSYH----- 471
 QY 67 ISVNGMRNNIKYTGSSP--NSVSSPTVSAITQLRNKLENGKPLSMSEOTS--LLIKIT 121
 DB 472 -----KKITVEKPYKCEGKAFNOSALLIKHKRIHTEKPYKCECGKTSKVST 523
 QY 122 EPLDENDYVLMATNG--FSSGSPFNNGGLAT---SPLVGPSAQSFMQLGYGMEAP 175
 DB 524 -----LTHKAIHAGEKPYKCEGKGTFLIVSTLTTHKAI---HAG--EKP 564
 QY 176 LIGFTPMNSNLSEVQVLIQVNTVSRQMDCKTSDISLKGYNMKDPCSGQEGQVT-- 233
 DB 565 Y-----KCEGCGAFS-----KPSILTHKVIHGEKPYKCEGKAFN 603
 QY 234 -SPNIPPVGLPVVSHNGATKSIIDYTLKVNBAKACLOSLTDSRQISIKKELRTL 292
 DB 604 MSNLT-----MEH---KRI--HTGEKPYKCEBGKSFSTFS----- 634
 QY 293 DLVTDKMLIENHSISTPSCQPKESFPGPILHQR-----YLCMNEEI 339
 DB 635 -VLTAKVI--HTGEKPYKCEGKAYKWSSTLSYHKIHTVEKPYK---CEGKAFNR 688
 QY 340 KAVLOPHENIVNKNAGVPYDNAL-----LLSVLSKGLTSPINPK----- 382
 DB 689 SALLIKHKI-----HTDEKPYKCEGKTSKSTLTTHKAIHAGEKPYKCEGKA 741
 QY 383 -DHMSVLYKAYVAMNM--EENSDELKLSIYAVLPOEVKWEQKRVYQYSNSRPSLER 439
 DB 742 FSKFSLTLTHKVIHTEKPYKCE-----ECGKAYNM-----PSTLS 777
 QY 440 TSKPLAPNSNPTTKOSLFRSPYKPMDSITSPSIALHNSVTSQDPPLTL--KSHPT 496
 DB 778 YHKIHTGEKPYKCEEC-----GKGFSMFSILTKHVIHTGEKPYKCEGKAFSWL 829
 QY 497 NIKAVDKLHSRS-----NTPSPMLSTSSKNSHS-----SYT 531
 DB 830 SVFSKRIKTHAGEKPYKCEACGKAYVTSIL--TKHVIHTGEKPYKCEGKAFNWS 885
 QY 532 PWSFSSEEL-----QAEPLD--LSLPKQMBEPKGIITKTKTKATSLNLDHNSVSSS 582
 DB 886 SNLMEKKIHTGETPYKCEBGCKAFSPWSSLTENKATTHAGEKPYKCEGKAFSWSRLL 945
 QY 583 EN-----SDEPLNLTPIKKEFSNSNNLDNKSNNPVFGMPPSAKPLVPLPQ----- 630
 DB 946 EHKATTHAGEBPYKCEGKAFNWSNL-----MEHKRIHTEKPYKCEBGCK 992
 QY 631 --SAPRPATFMPVQVSIPLGLRPYGLDOMSFLPHMAVYTPGAATFAMQGRK--YORK 687
 DB 993 SPTFSILTKHKVIHT--GEKPYK-CEECG---KAYWSSSTLSYHKIHTVEKPYKCE 1044
 QY 688 OGFGGLLDGAQDYMSGLDMDTSDSLRKKIKTRESGMVACDLCTFOKSSSLIRK 747
 DB 1045 ECGKGVV-----PSILAKHKVIHGEKPYKCEGKAYKMPSTLIRYK 1088
 QY 748 YEHTGRPHQCOICCKAFKHKHLIHSRLHSGEKPYCDCKGKRFSGSYSQH 802
 DB 1089 KIHTEGKPYKCEGKAFSTFSILTGHKVIHTEKPYKCEGKAFKAPSWLSVFSKH 1143
 RESULT 15
 ZF25 HUMAN STANDARD; PRT; 783 AA.
 ID ZF25 HUMAN
 AC Q9UT15;
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein ZFP25.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2004, 16:27:10 ; Search time 46 Seconds
(without alignments)
5295.688 Million cell updates/sec

Title: US-09-964-238-2

Perfect score: 4991
Sequence: 1 MLTQAGNRKFKTECGAF.....DGKMTKSDHEEDNEDMG 944

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_virus:*
- 12: sp_virulent:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4273	85.6	1214	13	Q91B60
2	4270	85.6	1213	13	Q91B60
3	1647.5	33.0	1104	4	Q8NB68
4	1646	33.0	306	11	Q8BSG9
5	1606.5	32.2	1154	4	Q13088
6	1449.5	29.0	1078	13	Q9DEH7
7	817.5	16.4	747	5	Q9VA39
8	817.5	16.4	1054	5	Q9VA40
9	757	15.2	650	5	Q94196
10	533.5	10.7	569	5	Q8MSQ8
11	365	7.3	189	11	Q8CSP2
12	343.5	6.9	914	11	Q8C827
13	341.5	6.8	949	11	Q35700
14	331.5	6.6	743	11	Q8R5D1
15	331.5	6.6	751	4	Q8TBA9
16	331.5	6.6	752	11	Q8BT00

17	328.5	6.6	569	4	Q9H9S6	Q9H9S6 homo sapien
18	328.5	6.6	701	4	Q9UG02	Q9UG02 homo sapien
19	324.5	6.5	819	4	Q8N946	Q8N946 homo sapien
20	322.5	6.5	688	4	Q8TA05	Q8TA05 homo sapien
21	321.5	6.4	616	4	Q8N9M3	Q8N9M3 homo sapien
22	319.5	6.4	1031	11	Q8CJ78	Q8CJ78 mus musculus
23	315	6.3	578	11	Q8BLA2	Q8BLA2 mus musculus
24	313.5	6.3	648	11	Q8K167	Q8K167 mus musculus
25	312.5	6.3	599	4	Q8TF20	Q8TF20 mus musculus
26	309	6.2	898	13	Q12958	Q12958 oryzias lat
27	306.5	6.1	1261	13	Q90ZM5	Q90ZM5 gallus gall
28	306	6.1	625	11	Q8BIN6	Q8BIN6 mus musculus
29	304.5	6.1	623	11	Q62510	Q62510 mus musculus
30	302.5	6.1	821	4	Q96JF6	Q96JF6 gallus gall
31	302.5	6.1	1173	13	Q907F3	Q907F3 gallus gall
32	299.5	6.0	1615	13	Q57415	Q57415 gallus gall
33	298.5	6.0	1186	11	Q08961	Q08961 rattus norv
34	297	6.0	468	4	Q8TF47	Q8TF47 homo sapien
35	297	6.0	738	4	Q9ULA1	Q9ULA1 homo sapien
36	297	6.0	871	4	Q8WTR7	Q8WTR7 homo sapien
37	296.5	5.9	637	4	Q96IR2	Q96IR2 homo sapien
38	296.5	5.9	1377	13	Q9DDN5	Q9DDN5 xenopus lae
39	296	5.9	786	4	Q8TDG8	Q8TDG8 homo sapien
40	295.5	5.9	769	4	Q81ZD3	Q81ZD3 homo sapien
41	295	5.9	648	4	Q81YB9	Q81YB9 homo sapien
42	294	5.9	914	4	Q9ULS9	Q9ULS9 homo sapien
43	292.5	5.9	714	4	Q8TE95	Q8TE95 homo sapien
44	291	5.8	873	6	Q9XSR1	Q9XSR1 canis famli
45	290.5	5.8	1350	13	Q91929	Q91929 xenopus lae

ALIGNMENTS

RESULT 1

ID Q91B60 PRELIMINARY; PRT; 1214 AA.

AC Q91B60; 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE Smad interaction protein 1.

GN SIPL.

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

OC Xenopodinae; Xenopus.

OX NCBI_TaxID=8335;

RN [1]

RP SEQUENCE FROM N.A.

RA Eisaki A., Kuroda H., Hayata T., Asashima M.;

RT "Xenopus laevis SIPL.";

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB038353; BAA94081.1; -

DR InterPro; IPR001356; Homeobox.

DR InterPro; IPR007087; Znf_C2H2.

DR Pfam; PF00096; Zf-C2H2; 8.

DR ProDom; PD000003; Znf_C2H2; 1.

DR SMART; SM00389; HOX; 1.

DR SMART; SM00355; Znf_C2H2; 7.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.

DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 6.

KW Metal-binding; Zinc; Zinc-finger.

SQ SEQUENCE 1214 AA; 136872 MW; 8D64868F652FB325 CRC64;

Query Match 85.6%; Score 4273; DB 13; Length 1214;
Best Local Similarity 86.1%; Pred. No. 3.6e-256;
Matches 815; Conservative 56; Mismatches 68; Indels 8; Gaps 5;

QY 1 MLTQAGNRKFKTECGAFYKHLKHLRIHSGEKYECPNCKKRSKSHSGSYSHSS 60
DB 272 MLTQAGNRKFKTECGAFYKHLKHLRIHSGEKYECPNCKKRSKSHSGSYSHSS 331

```

QY 61 KCCIGLISVNGRMNNIKTGSSPNSVSSSPTNSAITQLRNLKENGKPLSMSEQTGLKIK 120
DB 332 KCCIGLISVNGRMNNIKTGSSPNSVSSSPTNSAITQLRNLKENGKPLSMSEQTGLKIK 391
QY 121 TEPIDNDYKVLMAHFGSSGSPFNNGGLGATSPGCHPSAOSPMOHLGVGMEALPLGF 179
DB 392 TSLDNDYKVLMAASHAFNGAHPFNNGGLGATSPGCHPSAOSPMOHLGVGMEALPLGF 451
QY 180 PTMNSVLSEVQKVLQIVDNTVSRQKMDCTEDISKLGYMKDPCSQPEEGVTSFNIP 239
DB 452 PTMNSVLSEVQKVLQIVDNTVSRQKMDCTEDISKLGYMKDPCSQPEEGVTSFNIP 511
QY 240 VGLPVVSHNGATKSIIDYTLKYNKAKACLSLTTDSRQISNIKKKRLTLDLVTDK 299
DB 512 VGLPVVSHNGATKSIIDYTLKYNKAKACLSLTTDSRQISNIKKKRLTLDLVTDK 571
QY 300 MLENHSITPPFCQFCKESPPGIPPLHQRHRYLCNNKEIKAVLOPHENIVPKAVPND 359
DB 572 MLENHSITPPFCQFCKESPPGIPPLHQRHRYLCNNKEIKAVLOPHENIVPKAVPND 631
QY 360 NKALLSVLSEKGLTSPINPYKDHSVLYKAYYAMNENPDELKISLAVGLPOEFVKE 419
DB 632 KQALLSVLSEKGMSTPILYKDHSVLYKAYYAMNENPDELKISLAVGLPOEFVKE 691
QY 420 WFEQRVYQYANSRSPSLERTS--KPLAPNSPTTKDLSLPSVPKPMDSITSPSIAELH 477
DB 692 WFEQRVYQYANSRSPSLERTS--KPLAPNSPTTKDLSLPSVPKPMDSITSPSIAELH 749
QY 478 NSTSCDPLRLTKSHFTNIKAV-DKLDHSRNTSPPLNLSTSSKNSSSYTPNSFS 536
DB 750 NRSVNCDTPLRLTKSHFTNIKAV-DKLDHSRNTSPPLNLSTSSKNSSSYTPNSFS 809
QY 537 SEELQAEPLDLSPKQMRBPKGIATKNTKATSNILDNHVSASSSENDEPLNTFIKK 596
DB 810 SEELQAEPLDLSPKQMRBPKGIATKNTKATSNILDNHVSASSSENDEPLNTFIKK 869
QY 597 EFSNSNNLNDKSNPVFGNPPSAKPLVTPLPQSAFPATFMPVQTSIPGLRYPGDL 656
DB 870 EFSNSNNLNDKSNPVFGNPPSAKPLVTPLPQSAFPATFMPVQTSIPGLRYPGDL 927
QY 657 QMSFLPHMATYPTGAATPADMOQRKRYOKGFGODLDDGADYNSGLDMDTSDSCLS 716
DB 928 QMSFLPHMATYPTGAATPADMOQRKRYOKGFGODLDDGADYNSGLDMDTSDSCLS 987
QY 717 RKIKITTESGMVACDCTFKOKSSSLRHKYHTGKRPQCOICCKAFGKHLLHSHR 776
DB 988 RKIKITTESGMVACDCTFKOKSSSLRHKYHTGKRPQCOICCKAFGKHLLHSHR 1047
QY 777 LHSGEKPYQCDKCGKRFSGSGSYSOHNNHRYSCKBAEREAERAREKHLGPTLL 836
DB 1048 LHSGEKPYQCDKCGKRFSGSGSYSOHNNHRYSCKBAEREAERAREKHLGPTLL 1107
QY 837 MNNAVYQSTIPQGYSDSEERESMPROGESKEHEKEBEGYGLRRDGESEEBEES 896
DB 1108 MNNAVYQSTIPQGYSDSEERESMPROGESKEHEKEBEGYGLRRDGESEEBEES 1167
QY 897 ENKSMPTDPTIRDEEETGDSHMDSDSEDGMETKSHDEEDNMDGM 943
DB 1168 ENKSMPTDPTIRDEEETGDSHMDSDSEDGMETKSHDEEDNMDGM 1214

```

RESULT 2

```

Q918P1 PRELIMINARY; PRT; 1213 AA.
AC 0918P1;
DT 01-OCT-2000 (TREMBlurel. 15, Created)
DT 01-OCT-2000 (TREMBlurel. 15, Last sequence update)
DE 01-MAR-2003 (TREMBlurel. 23, Last annotation update)
GN Smad-interacting protein 1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

```

```

OC Xenopodinae; Xenopus.
OX NCBI_TaxId=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20302561; PubMed=10842070;
RA van Grunsven L.A., Papin C., Avalosse B., Opdecamp K., Huybregt D.,
RT "XSLP1, a Xenopus zinc finger/homeodomain-encoding gene highly
RT expressed during early neural development.";
RL Mech. Dev. 94:189-193 (2000).
DR EMBL; AF237679; AAF81689.1; -.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 8.
DR Prodom; PD000003; Znf_C2H2; 1.
DR SMART; SM00355; ZNF_C2H2; 7.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 6.
KW Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 1213 AA; 136775 MW; E0AB78D294A2877B CRC64;

```

```

Query Match 85.6%; Score 4270; DB 13; Length 1213;
Best Local Similarity 86.0%; Pred. No. 5,6e-256;
Matches 814; Conservative 57; Mismatches 68; Indels 8; Gaps 5;

```

```

QY 1 MTOGAGNRKFKTECGKAFKTKHLKEHLRHSGBKPYECNCKRFSHSGSYSHSS 60
DB 271 MTOGAGNRKFKTECGKAFKTKHLKEHLRHSGBKPYECNCKRFSHSGSYSHSS 330
QY 61 KCCIGLISVNGRMNNIKTGSSPNSVSSSPTNSAITQLRNLKENGKPLSMSEQTGLKIK 120
DB 331 KCCIGLISVNGRMNNIKTGSSPNSVSSSPTNSAITQLRNLKENGKPLSMSEQTGLKIK 390
QY 121 TEPIDNDYKVLMAHFGSSGSPFNNGGLGATSPGCHPSAOSPMOHLGVGMEALPLGF 179
DB 391 TSLDNDYKVLMAASHAFNGAHPFNNGGLGATSPGCHPSAOSPMOHLGVGMEALPLGF 450
QY 180 PTMNSVLSEVQKVLQIVDNTVSRQKMDCTEDISKLGYMKDPCSQPEEGVTSFNIP 239
DB 451 PTMNSVLSEVQKVLQIVDNTVSRQKMDCTEDISKLGYMKDPCSQPEEGVTSFNIP 510
QY 240 VGLPVVSHNGATKSIIDYTLKYNKAKACLSLTTDSRQISNIKKKRLTLDLVTDK 299
DB 511 VGLPVVSHNGATKSIIDYTLKYNKAKACLSLTTDSRQISNIKKKRLTLDLVTDK 570
QY 300 MLENHSITPPFCQFCKESPPGIPPLHQRHRYLCNNKEIKAVLOPHENIVPKAVPND 359
DB 571 MLENHSITPPFCQFCKESPPGIPPLHQRHRYLCNNKEIKAVLOPHENIVPKAVPND 630
QY 360 NKALLSVLSEKGLTSPINPYKDHSVLYKAYYAMNENPDELKISLAVGLPOEFVKE 419
DB 631 KQALLSVLSEKGMSTPILYKDHSVLYKAYYAMNENPDELKISLAVGLPOEFVKE 690
QY 420 WFEQRVYQYANSRSPSLERTS--LAPNSPTTKDLSLPSVPKPMDSITSPSIAELH 477
DB 692 WFEQRVYQYANSRSPSLERTS--LAPNSPTTKDLSLPSVPKPMDSITSPSIAELH 748
QY 478 NSTSCDPLRLTKSHFTNIKAV-DKLDHSRNTSPPLNLSTSSKNSSSYTPNSFS 536
DB 749 NRSVNCDTPLRLTKSHFTNIKAV-DKLDHSRNTSPPLNLSTSSKNSSSYTPNSFS 808
QY 537 SEELQAEPLDLSPKQMRBPKGIATKNTKATSNILDNHVSASSSENDEPLNTFIKK 596
DB 809 SEELQAEPLDLSPKQMRBPKGIATKNTKATSNILDNHVSASSSENDEPLNTFIKK 868
QY 597 EFSNSNNLNDKSNPVFGNPPSAKPLVTPLPQSAFPATFMPVQTSIPGLRYPGDL 656
DB 869 EFSNSNNLNDKSNPVFGNPPSAKPLVTPLPQSAFPATFMPVQTSIPGLRYPGDL 926
QY 657 QMSFLPHMATYPTGAATPADMOQRKRYOKGFGODLDDGADYNSGLDMDTSDSCLS 716
DB 927 QMSFLPHMATYPTGAATPADMOQRKRYOKGFGODLDDGADYNSGLDMDTSDSCLS 986

```

Db	386	-----SINLSDIQNVLKVADVGNVIRQVLE---NNQANLASKEQETINASPIQGGGHSV	436
Qy	236	NIPPVGLPVSNHNGATKSTIDYILEKKNNAKCLQSITTSRRQISINIKKEKATLIDLIV	295
Db	437	-ISAIISPLVDQOGTTKIIINYSLEQSSQOQVVPQNKKENPAPVATNSCKSKXKLPEDILTVK	495
Qy	296	TD-DKMIENH-SISTPSCQFCCKESFGPI-PLIHOHERYLCLCKNNEELKAVLQPHENIVPN	352
Db	496	SEKDKSEGGVNDSTCLLDDC---PGDINALPELKYH-----DLKQPIQP-----PP	540
Qy	353	KAGVVDNKKALLSSVLSSEKGLTSPINPYDHMSVLKAYVAMNNEPPNSDELKTSIAGVL	412
Db	541	LPAABAEKPESSVSSATGDNLSPPSQPPLKNLISLLKAYVALNAQPSAEELSKTADSVNL	600
Qy	413	POEFVYKEMFQQRKYQYQNSRSRSLERTSKPLAPNSPTTKDLSLPPSPVYPMNISTSPS	472
Db	601	PLVVYKMFQKMAQGIS-----VQSSSPSPSEPKV-----NIPAKNDQPOSAN	646
Qy	473	IAELHNSVTCDDPPLRLTKSSHFTNIAKAVDKLIDSRNTPSPNLSTSSKNSSSYTP	532
Db	647	ANRPQDSTVNLQSPDKNTNS---PVLPGVSTNRSRSTSPSLANLSSNGTQGYLYTA	703
Qy	533	NSFSSEELQAPLIDLSPKQMRPKGIIATKNKTATSNLDHNSVSSSSSEN-----DE	587
Db	704	EG-AQEEPVEPDLDSLPKQGG-----LLERSTITSVYQNSVYSVQEE	746
Qy	588	PLNLTPFKKPFNSNNLDKNSNPVFGMPPSAKPLTLPPLPQSAFPATMPVQTSIP	647
Db	747	PLNLSCAKKRPQKQSCVTD---SEVVAVIIPPSAPINIAIPLTYAQLPTTIVALADQNSVP	804
Qy	648	GLRPYPGLDQMSFLPHMAVYTPGCAATFADMQQRKYQKQKQFGDLDGAQDYMSGDD	707
Db	805	CLALANKKTILLPQAYITSTVSP-AVQEPPLKYIQPNGNDERQDITSBEQSVNED	863
Qy	708	MTDSDCLSRKKIKTRESGYACDLCDRTQKSSSLIRHKYERTGKRPQOICCKAKFK	767
Db	864	QNSDSDPYPKKKMRKTENGAYACDLCDKIFQKSSSLIRHKYERTGKRPHEGICCKAKFK	923
Qy	768	KHLIIEHSLHSGEKPYQCDKCGRPFHSGSYQNMHNRYSYCKREAEERPAAREARK	827
Db	924	KHLIIEHRLHSGEKPYQCDKCGRPFHSGSYQNMHNRYSYCKREAEERDSTQE---	979
Qy	828	GHLGPTLLANRAYLQITP-QGYSDSEESMPPR--DGESEKEHEKEGE-----GYK	879
Db	980	EAGP-EILSNENHVGAAASPSQG--DSDEKESLIREEDDEDEKKEEBEDKMELOGEKE	1035
Qy	880	LRRRDDEEEEBEESSEKSMQDTPETIRDEETGHSMDSSD-----GKMET	930
Db	1036	CEKPGQGEEGEEEBEVEEBEVE-EAENEGEAEKTEGLMKDRAEQASLSGQKGESE	1094
Qy	931	KSDHEEDN 938	
Db	1095	QVSEKTN 1102	
RESULT 4			
Q8BSG9			
ID	Q8BSG9	PRELIMINARY;	PRT; 306 AA.
AC	Q8BSG9;		
DT	01-MAR-2003 (Tremblere). 23, Created)		
DT	01-MAR-2003 (Tremblere). 23, Last sequence update)		
DT	01-MAR-2003 (Tremblere). 23, Last annotation update)		
DE	zinc finger homeobox 1b.		
OS	Mus musculus (Mouse)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
OX	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=Mesonephros;		
RX	MEDLINE=22354683; PubMed=12466851;		
RA	THE FANTOM Consortium,		

RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK032970; BAC28104.1; -.
SQ SEQUENCE 306 AA; 35506 MW; 730827A3D350464C CRC64;

Query Match 33.0%; Score 1646; DB 11; Length 306;
Best Local Similarity 99.7%; Pred. No. 3.2e-94;
Matches 304; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 639 MPVQVSIIGRLRYPGLDQMSFLPHMAATYTPGAATFADMQQRKYQKQGFQGDLLDA 638
DB 1 MPVQVSIIGRLRYPGLDQMSFLPHMAATYTPGAATFADMQQRKYQKQGFQGDLLDA 60
QY 699 ODYMSGLDMDTSDSCLSRRKIKTTESGMYACDLCKTFQKSSSLRHKYEHTGKPHQC 758
DB 61 QDVMGLDMDTSDSCLSRRKIKTTESGMYACDLCKTFQKSSSLRHKYEHTGKPHQC 120
QY 759 QICKKAFKXGHLIEHSRLHSGEKPYQCDKCGKRFSGSGYSQHMNRYSYCKREAERE 818
DB 121 QICKKAFKXGHLIEHSRLHSGEKPYQCDKCGKRFSGSGYSQHMNRYSYCKREAERE 180
QY 819 AAREAREKCHGPTLLMRAVLOSTTPQGYSDSEEREMPRDGSSEKHEKEGEGYG 878
DB 181 AAREAREKCHGPTLLMRAVLOSTTPQGYSDSEEREMPRDGSSEKHEKEGEGYG 240
QY 879 KLRRRDGEDEEESEENKSMPTPETRDEEETGDSMDSDSEDGKMETSDHEEDN 938
DB 241 KLRRRDGEDEEESEENKSMPTPETRDEEETGDSMDSDSEDGKMETSDHEEDN 300
QY 939 MEDGM 943
DB 301 MEDGM 305

RESULT 5
ID Q13088 PRELIMINARY; PRT; 1154 AA.
AC Q13088;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE ZEB (Fragment).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Genetica T., Ruzsinszky D., Kadesch T.;
RT "Displacement of an E-box-binding repressor by basic-helix-loop-helix
RT proteins: Implications for B-cell specificity of the immunoglobulin
RT heavy-chain enhancer.";
RL Mol. Cell. Biol. 15:0-0(1995).
DR EMBL; U19969; AAA62155.1; -.
DR TRANSFAC; T00625; -.
DR TRANSFAC; T02324; -.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; Zf-C2H2; 7.
DR ProDom; PD000010; Homeobox; 1.
DR ProDom; PD000003; Znf_C2H2; 2.
DR SMART; SM00389; HOK; 1.
DR SMART; SM00355; Znf_C2H2; 7.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 6.
KW Metal-binding; Zinc; Zinc-finger.
FT NON TER 1
SQ SEQUENCE 1154 AA; 126924 MW; 1A4DAC7851909113 CRC64;

Query Match 32.2%; Score 1606.5; DB 4; Length 1154;
Best Local Similarity 41.1%; Pred. No. 5.2e-91;

Matches 399; Conservative 142; Mismatches 314; Indels 115; Gaps 32;

QY 2 LTQAGNRKFKCTEGCAFKYGHLEKHLIRISGEKPYECPCKRFSHSGSYSHSSK 61
DB 265 VTQSGNRKFKCTEGCAFKYGHLEKHLIRISGEKPYECPCKRFSHSGSYSHSSK 324
QY 62 KICGLISVNGRMRNITG--SSPNSVSSPTNSAITQLRNLKENGKPLSMEOGLKAI 119
DB 325 KICSLIPVNGRMRNITG--SSPNSVSSPTNSAITQLRNLKENGKPLSMEOGLKAI 380
QY 120 KTEPLDFNDYKILMTHTGSGSSPFMNGGLATSPGLGVHPSAQSPQGH--GVCMEAPL 176
DB 381 KTEPLDY--EFKPLVVASGINCSTPLQNGVFTGGGFLQATSSFGQVAVVLTGGLVSI 439
QY 177 LGEPITNSMSEVOKVQL--VDNTVSROQMDCKTEDISKLGYHMKDPCSOBEQGVSP 235
DB 440 -----SILSLDIQNLKAVDGNVIRYLE--NNQANLAKGEOTINASTIQCGHGV 490
QY 236 NIPVGLPVSHNGATKSIIDYTLKVNKAACLOSLTDSRQISNIKKELRTIDLV 295
DB 491 -ISAISLPLVDQDTTKIINYSLEQPSQLQVVPQWLKKENPVATNSCKSEKLPEDLTVK 549
QY 296 TD--DKMIENH--SISTPFGQFCKESPPI--PLHNERLYCKMNEIKAVLDPHEIVN 352
DB 550 SEKDYSFEGGVNDSTCLCDDC---PGDINALPELKH--DLKOPTP---PP 594
QY 353 KAGVPVNDKALLLSVLSSEKGLTSPINPKDHMSVLKAYYANMNEBDELKISAVGL 412
DB 595 LPAAEKREKRESSVSAITGNGNISPSQPLKNLISLKAYVALNAQSAEELKIDSVNL 654
QY 413 PQEFVKEPQKRYOYNSRSPSLERTSKPLAPNSNPTTKOSLLPRSGVKKPMDSITSPS 472
DB 655 PLDVVAKMPEKMQAGQIS-----VQSESEPSPEPKV-----NIPAKNNDQPOSAN 700
QY 473 IAEIHSVTSOCPRLTSSHTFNIAKVDKLDHSNSNTPSLNSTSKSHSSSYRP 532
DB 701 ANEPQSTVNLQSLPKMTNS---PVLPGSTTNGSRSTSPSLNLSSRNTQGYLVA 757
QY 533 NSFSSELOAEPDLDSLPRQMR-----PKGIATNKKTYATSINLDHNSVSSSENS 585
DB 758 EG--AGEEPQVPEPLDSLPRQGELEERILPLRQCLFCPRT----- 797
QY 586 DEPLNTITFKKPSNSNNDKSNPFVFGMPPSAKPLTPLPQSAFPAPATFMPVQTS 645
DB 798 ---LELVLAKEKPKQKSCVTD--SEPVVAVIPPSANPINIAIPVVAQLPTVAIDQNS 852
QY 646 IGLRYPGLDQMSFLPHMAATYTPGAATFADMQQRKYQKQGFQGDLLDGAQDVMGL 705
DB 853 VPCLRALANKOTILIPVAVYISTVSP-AVOEPLKVIQPNGNODERQDTSBGVSNV 911
QY 706 DDMTSDSCLSRRKIKTTESGMYACDLCKTFQKSSSLRHKYEHTGKPHQCQICKKAF 765
DB 912 EQNDSDSIFPKKKRKKTENGAYACDLCKITQKSSSLRHKYEHTGKPHQCICKKAF 971
QY 766 KKHKHLIEHSRLHSGEKPYQCDKCGKRFSGSGYSQHMNRYSYCKREAEREBAAREAR 825
DB 972 KKHKHLIEHSRLHSGEKPYQCDKCGKRFSGSGYSQHMNRYSYCKREAEREBAAREAR 1029
QY 826 EKCHGAGPTLLMRAVLOSTTPQGYSDSEEREMPRDGSSEKHEKEGEGY-----GY 877
DB 1030 ---EAGP--BILSNHVGARAPSPQ--DSDEBSLTFREDESEKEEEDKEMELOEE 1083
QY 878 GKLRRRDGEDEEESEENKSMPTPETRDEEETGDSMDSDSEDGKMETSDHEEDN-----GKM 928
DB 1084 KECEKQGEDEEESEENKSMPTPETRDEEETGDSMDSDSEDGKMETSDHEEDN-----GKM 928
QY 929 ETKSDHEEDN 938
DB 1143 SEQVSEKTN 1152

RESULT 6
Q9DEH7

ID Q9DEH7 PRELIMINARY; PRT; 1078 AA.
AC Q9DEH7;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Kheper.
GN ZFH1 OR KHEPER.
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
NCBI_Taxid=7955;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20541661; PubMed=11087624;
RA Murooka O., Ichikawa H., Shi H., Okumura S., Taira E., Higuchi H.,
Hirano T., Hibi M., Miki N.;
RT "Kheper, a Novel ZFH/deltaEF1 Family Member, Regulates the Development
of the Neuroectoderm of Zebrafish (Danio rerio).";
RL Dev. Biol. 228:29-40(2000).
DR EMBL: AB016799; BAB18865.1; -.
DR ZFIN: ZDB-GENE-010621-1; zfhx1.
DR InterPro: IPR001356; Homeobox.
DR InterPro: IPR007087; Znf_C2H2.
DR Pfam: PF00096; homeobox_1.
DR Pfam: PF00096; zfc_C2H2; 7.
DR Prodom: PD000010; Homeobox; 1.
DR Prodom: PD000003; Znf_C2H2; 2.
DR SMART: SM00389; Hox; 1.
DR SMART: SM00355; Znf_C2H2; 7.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 7.
KW Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 1078 AA; 117844 MW; B3D236A69D1235BC CRC64;

Query Match 29.0%; Score 1449.5; DB 13; Length 1078;
Best Local Similarity 37.6%; Pred. No. 2.6e-81;
Matches 376; Conservative 126; Mismatches 285; Indels 213; Gaps 31;

QY 5 GAGNRKFKCTEGCKAFKTKHLEHLRIHSGEKYPCEPCNCKRFSHSGSYSHSSKCI 64
DB 203 GGGNKKFKCTEGCKAFKTKHLEHLRIHSGEKYPCEPCNCKRFSHSGSYSHSSKCI 262
QY 65 GLISVNGMRNNIKTG-----SSPNSSVSSPTSAITQLNKLENGKPLSMSTQGL 117
DB 263 GLISVNGMRNNIKTG-----SSPNSSVSSPTSAITQLNKLENGKPLSMSTQGL 117
QY 118 KIKTEPLDFNDYKVMATHTGFSGSPFNGGIGATSPGVHPSAOSPQHLGVMEAPL 177
DB 319 QIKSEPLDY-EYKPVVAPASAGVGMFGG-----AAPLQAGAVAVLPTV 365
QY 178 GEPPT-MNSNLSFVQYLOI-VNNTVSROKMDCKTEDISKLYHKMDPCSGPEQGV--- 232
DB 366 GLVSPISINLGLQNVNKKVAVDGVNIRQVLES-----TQAK-----CQAGTGIYGA 412
QY 233 -----TSPNTPPVGLPVVSHNGATKSTIIDYLEKVNKAKAGLSLTDSRROINIKKE 286
DB 413 GGIAATQOQVIOATISPLIDQGNAKIFNYSLDP-SQQAVALQS-----PKKE 460
QY 287 KLRITLIDLVTDKMLNHSISTPSCOFCKESPFGPIPIHQHERYLCKNNEIKAVLOPH 346
DB 461 PLGSTEVCCKGKGLKEDDLTVK-----NRD----- 485
QY 347 ENIVNKAQVFNQKALLISVLS-----KGLTSPINPYKDHMS 386
DB 486 -----KTLTYDEKSMNLNDILKRCGDEHRINGKNLEKMDLEGLCPGQPLKNLIS 539
QY 387 VLKAYVAMNENBELKISIAVGLPOFVKEVFEORQVYOS-NSRSPSLERTSKPLA 445
DB 540 LKAFALANNEPTKELAKISISVSLPAVVKWKEKMOGLGISNDPSSPGHEE----- 593
QY 446 PMSNPTTQDILPRSPVVKEMDITSPSIAELHNSVTSQDPLRLTKYSHFTNIKAVDKLD 505

DB 594 EOTTPVDDGTGKASPKEDDLQOWN-----SEKOEERCCSPAGIAAS-----VNGIESVP 645
QY 506 HSRNSPTPLNLSNSSKSHSSSYTPNSFSSEELQAEPDLPLSKQKREPKGIATQNK 565
DB 646 AS-----PSPNLNLSAG-----PVPARTLEGBEPLDLSL-----KSA 679
QY 566 TKATISINLDHNSVSSSSNSDEPLNLTPIKKEFSNNLNDKSNPNVPGMPPFSKAP- 622
DB 680 TTASVASHANTVYSAQ-----EPNLITCTKEL-----LSNASTNAALYASQPSANPLNI 731
QY 623 LYTPLPPOSAPPATMPPVOTSTIPGLRPYPGLDQMSFLPHNAYYPFGATFADMOOR 682
DB 732 MTTQLPPLVATD-----GQAQCLRALTTTKQTILIPOLTVSYTTTSSSPAGTDTPQ 784
QY 683 K-YRKQSGFDLIDGADVWMSGDDMDSDSCSRKIKKTESGMVACDICTKFOKSS 741
DB 785 KNLHVNQIKKEKQDMSGSEATSLIEQTDSPGPKKKMKRTESGLVACDICTKFOKSS 844
QY 742 SLNRKHYHTGKRPHQCIQCKAFKHLIHSRLHSGEKPYQCDKCKRFSHSGYSQ 801
DB 845 SLNRKHYHTGKRPHEGICSKAFKHLIHLRLHSGEKPYQCDKCKRFSHSGYSQ 904
QY 802 HNNRYSYCKRBAER-----EAEERAREKHLGPTLLMNRAYLOSITPGYSSE 855
DB 905 HNNRYSYCKRBAER-----EAEERAREKHLGPTLLMNRAYLOSITPGYSSE 855
QY 856 RESMPRDESEKHE-----KEGREGYCKLRRDGDSEEESESEKSS 900
DB 955 KGSSTRDSESEBDMGSLVDEDEIQVVKIGGEDDDQADBDQRMERBEGDGBSG 1014
QY 901 MDTPETIRDEETGDSHSDSDSEDKMETKSDHEEDNME 940
DB 1015 ETLKVVWQDCEBDENEDEQTSBDADT-SDHTEEMTE 1053

RESULT 7
QYVA39 PRELIMINARY; PRT; 747 AA.
AC QYVA39;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE ZFH1 protein.
GN ZFH1 OR CG1322.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Mecoptera; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_Taxid=7227;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Amandides P.G., Scherer S.B., Li P.A., Hoskins R.A., Galle R.F.,
Amanatides P.G., Scherer S.B., Li P.A., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champagne M., Pfeiffer B.D.,
Wan K.H., Doyle J., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
Abrill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J., Brokstein P., Broctier P.,
Burris K.C., Busam D.A., Butler H., Cadenot L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durlin K.J., Evangelista C.C., Ferreira C., Ferreira S., Fleischmann W.,
Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan F., Harris W.,
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Siden-Kimios I., Simpson M., Skueci M.P., Smith T.,
 Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
 Svrtak R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weinstein J.,
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.W., Venter J.C.,
 "The genome sequence of *Drosophila melanogaster*.";
 Science 287:2185-2195 (2000).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL: AB003775; AAF57083.1; -.
 DR HSSP: P08153; 1ZPD.
 DR Flybase: FBgn004606; zfh1.
 DR InterPro: IPR001356; Homeobox.
 DR InterPro: IPR007087; Znf_C2H2.
 DR Pfam: PF00046; homeobox_1.
 DR Pfam: PF00096; zf-C2H2; 9.
 DR Prodom: PD000010; Homeobox; 1.
 DR Prodom: PD000003; Znf_C2H2; 2.
 DR SMART: SM00389; HOX; 1.
 DR SMART: SM00355; Znf_C2H2; 9.
 DR PROSITE: PS00027; HOMEBOX 1; 1.
 DR PROSITE: PS00071; HOMEBOX 2; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 6.
 DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 9.
 DR DNA-binding; Homeobox; Metal-binding; Nuclear protein; Zinc;
 Zinc-finger.
 SQ SEQUENCE 1054 AA; 116597 MW; 5189AB214AB5BB CRC64;

Query Match 16.4%; Score 817.5; DB 5; Length 1054;

Best Local Similarity 29.1%; Pred. No. 3.5e-42; Indels 253; Gaps 33;

Matches 255; Conservative 101; Mismatches 268; Indels 253; Gaps 33;

9 RKFKTECGKAFKYLKHLKHLRIHSGEPEYCPNCKRFSHSGSYSHISKCGIGIS 68
 353 RKFKTECGKAFKYLKHLKHLRIHSGEPEYCPNCKRFSHSGSYSHISKCGIGIS 409
 69 VNGRRNN-----IKTGSSPNSVSSPTNSAITQLRNKLEN-----GKPLSMSEQTGLIKI 119
 410 MGLKLNRRALKRLKLEKSPGSSASRSRSPDHGKGLPEQSLPLGLPPMS----- 461
 120 KTEPLDPNDYKYLMAITHGSSGSPMNGGLATSP-LGVHPSAQSPMHLGVMEAPLIG 178
 462 -----YFASDAQVGGSAAPAPPPPPHPPV-----NMAALLA 493
 179 FP-----TMSNLS-----EYQKVLQIVDNTVSRQKMDCKTDSIKLGYHMKDPSOPE 228
 494 FPHNMAAAGLIDPRVHPISIORLQL--SAAGQOOREEREEOQKOOH--DEETPD 548
 229 EOGVTSPIIPVGLPVSHNGATKSIIDYTLKVNBAKACLOSLTDSRRQISNIKEKL 288
 549 EPKLVMDIEEPKTEKMAPTEATATPIKRESEEA-----SPPEGRSSQALIKEO 603
 289 RTLILVTDKMEIENS--ISTPPSCQPKSPFPIPLHCHERYLCKMNEITAVIOP 345
 604 EPLVAERQTPVEEHAPVHAADLRCSRCQFHPHPELVQHEKVLGCL--IKEELQ 660
 346 HENIIPNKAQVGV-----DNKALLSVLSKSG-----LNSPINPYKDHNSVKA 390
 661 H--FOOQQTSTALASASEDEDEBMDVEEPPROESGRKRVKRTAIN--EEQOOLQ 716
 391 YYANMPEPNDLKLISIAVGLPOEFVKEMPE-----QKVVYQYNSRS----- 434
 717 HYSILAPSRBDFPMIAARLQIDPRVYQWFGNNRSREKMQSPQNNQAGAAMPPIID 776
 435 -PSLEKTSKPLAPNSNPTTKOSLIPSPVYKPMDSITSPSIALHNSVTSQDPLALTYSS 493
 777 QASLTREDOPL-----DLSVKRDPILTPKSESSPPYIA-----PP-----SG 812

494 HFTNIKAVD-KLDHSRNTPEPLNLSTSSKNSHSSYTPNSFSSSEELQAP----- 544
 813 EALNDEALINLSRKSTSSMSBPASISPSAALVFGAAPSPSPSGOLDSTPRSGQAPG 872
 545 -----LDSLRLKQ-----MREKGIATKTKATKASINI--DHNSVSSSENSDEPLNLTPI 594
 873 LPPVYLPPSLPMEALFKRR-FCGDFAS--NHAIIMSILKLPDRCSTLSFGGSB----- 922
 595 KKEPNSNNLNKSNPPVFGNPPSAKPLTPLPQSAFPATMPPPQTSIPGLRPYGP 654
 923 KRSWDD-----SRISHDEFAGVLTMP----- 947
 655 LDQMSFLPHMAYTPGATPADMQORRYOKQFOGDLIDGAQDYMSGLDMDTSDSC 714
 948 -----KPRGKAETHIGHAD-----PDLR----- 966
 715 LSRKKIKTESGATACIDYTPQKSSLLRHKTEHTGRPHQOCCKKAFKHLIETH 774
 967 -----YVCDQCDKAFKQSSILAHKTEHSGQRPYQCIECPKAFKHLIETH 1013
 775 SRLHSGEPEYCPNCKRFSHSGSYSHISKCGIGIS 811
 1014 KRLHSGEPEYCPNCKRFSHSGSYSHISKCGIGIS 1050

RESULT 9

ID 094196 PRELIMINARY; PRT; 650 AA.

AC 094196; 01-FEB-1997 (TREMBLrel. 02, Created)

DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE Zinc finger plus homeodomain, axon guidance protein 1.

OS *Caenorhabditis elegans*.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;

OC Rhabditidae; Peloderinae; *Caenorhabditis*.

OX NCBI_Taxid=6239;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RX MEDLINE=9906613; PubMed=9851916;

RA Waterston R.;

RT "Genome sequence of the nematode *C. elegans*: a platform for

investigating biology. The *C. elegans* Sequencing Consortium.";

RL Science 282:2012-2018 (1998).

RN (2)

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Nelson J., Mohlmann P.;

RT "The sequence of *C. elegans* cosmid P28P9.";

RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.

RN (3)

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Waterston R.;

RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: U70850; AB09122.2; -.

KW Homeobox; DNA-binding; Nuclear protein.

SQ SEQUENCE 650 AA; 72596 MW; A40DA56CE7A95F1 CRC64;

Query Match 15.2%; Score 757; DB 5; Length 650;

Best Local Similarity 25.8%; Pred. No. 1e-38; Indels 366; Gaps 28;

Matches 229; Conservative 102; Mismatches 191; Indels 366; Gaps 28;

9 RKFKTECGKAFKYLKHLKHLRIHSGEPEYCPNCKRFSHSGSYSHISKCGIGIS 68
 22 RKFKTECGKAFKYLKHLKHLRIHSGEPEYCPNCKRFSHSGSYSHISKCGIGIS 77
 69 VNGRRNNIKTGSSPNSVSSPTNSAITQLRNKLENKPLSMSEQTGLIKITEPLDND 128
 78 -----QOASPSWVT-----PPNP 90


```

QY 129 YKVLMTAGFSGSSPPMNGGLATGATGVLGHPAQAQSPMOHLGVGMEAPLLGF-PTMNSLNS 187
DB 91 YQJLMM-----YRNIMLQLOTPQVSLFSTANNN 119
QY 188 EVQKVLQIVDNVYSRQKMDCKTEDISKAGYHMKDCSQPBEQGVTSNIPVPGLVVGH 247
DB 120 DVMSLQ-----ANLFOSLNGTSPPTQESAPASPEPKLEVDEBEVS- 164
QY 248 NGATKSIIDYLEKVNKAACLOSLTTSRROISNIKEKRLTLDLVTDKMIENHSIS 307
DB 165 -----SEVTE-VKI-EVKTEDSVPF-ESIT 187
QY 308 TPFSQCFCKESFPGPPIPLHQHERYLCKMNEIKAVLQPHENIVPKAGVFVDNKA----- 362
DB 188 PAVSMGL-----SPAP-----EQNGNEBMNNGSGSDCKSPDW 222
QY 363 -LLLSVLSBKGLTSPINPYKDMSVLKAYAMNMPNSDELLKISIVGLPQEPFKEMF 421
DB 223 PLASRSPIND-----SCVAVLQNHFKKNPFPKSKLESAVAEOIGVNRVVQVWF 271
QY 422 EQRKVQYSNRSRPSLERTSKPLAPNSNPTTKDSSLPRSPVKMDSTSPSIAELHNSVT 481
DB 272 QNTRAKERSNRPLSPRGSVASAAAA-----AATSTTWQ----- 307
QY 482 SCDPPLRLTK--SSHFTNIKAVDKLHRSRNTSPPLNLSSTSKNSHSSSYTPNSPSSSE 539
DB 308 -TPVOLMAAMASQFSN-----GNNSLTASQDERNNENTDEVMHDHC 347
QY 540 L---QAEPLDLSLPKQMRPK-----GIATKTKKARS 570
DB 348 LKDGKTPDLDLTSTDTDEPMSPEKLIQGLDQGVLOELLQAQNGFVNTDEDEBEK 407
QY 571 INLDHNSVSSSEN-----SDEPLNLTFKKFSNNLND----- 606
DB 408 IKABESPVSQSSSIMPSTFGVPSLIDSLSLVLEKALDVST--DFSQKPFDSIGR 465
QY 607 -----KSNMPVFGNPPSAKPLVPLPPQSAFPAPATMPVQTSIPGLRPYGL 655
DB 466 EDIRAVRYVVIYSNKOIF-----EKLYYP----- 489
QY 656 DQMSFLPMAYTPTGAATPADMOQRKRYKRGQFGODLLDGAQDYMSGLDMDTSDCL 715
DB 490 -----PRYALT-----NNFQDMQOK-----SSBDA--SSLCS 515
QY 716 SRKTI-----KTTESGYACDLCDKTFQKSSSLRHKYEHTGKPHOCQICKKAFKX 768
DB 516 NESKLLKFTPTPLKEBGLFSCQCDKVFQKOSSLARHKYEHSGQRPYKCDICEKAFKX 575
QY 769 HHLIHSRLHSGEKPYQCDKGRFSGSGYSQHMNRYSYCKRAE 816
DB 576 HHLTEHRLHSGEKPFQCDKLRFSHSGSYQHMNRYSYCKRPYRQ 623

RESULT 10
Q8MSQ8 PRELIMINARY; PRT; 569 AA.
AC Q8MSQ8;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE SD06902P.
GN ZFH1 OR CG1322.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota, Metazoa, Arthropoda, Hexapoda, Insecta, Pterygota;
OC Neoptera, Endopterygota, Diptera, Brachycera, Muscomorpha;
OC Ephydroidea, Drosophilidae, Drosophila.
OX NCBI TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,

```

```

RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragae V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celinker S.;
RL Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AY118654; AAM50023.1; -
DR FlyBase; FBgn004606; zfh1.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00046; homeobox.1.
DR Pfam; PF00096; zf-C2H2; 1.
DR Prodom; PD000010; Homeobox; 1.
DR Prodom; PD000003; Znf_C2H2; 1.
DR SMART; SM00389; HOX; 1.
DR SMART; SM00355; Znf_C2H2; 4.
DR PROSITE; PS00027; HOMEBOX 1; 1.
DR PROSITE; PS00071; HOMEBOX 2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 4.
KW Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 569 AA; 63968 MW; DC7141EPF2618B1 CRC64;

```

Query Match 10.7%; Score 533.5; DB 5; Length 569;
 Best Local Similarity 27.9%; Pred. No. 68-25;
 Matches 172; Conservative 68; Mismatches 187; Indels 189; Gaps 23;

```

QY 250 ATKSIIDYLEKVNKAACLOSLTTSRROISNIKEKRLTLDLVTDKMIENHS---I 306
DB 85 ATEAATPIKRBSREA-----SPDESYSNSQAIKQEDPLNVAERQTPVEBHAUVGH 139
QY 307 TPFSQCFCKESFPGPPIPLHQHERYLCKMNEIKAVLQPHENIVPKAGVF----- 358
DB 140 SADLRCSRSCKFNPHTLVQHEKVLGCL--IKELEQH--FQOQQAISFALASASED 194
QY 359 --DNKALLSVLSEKQ-----LTSPIPKDMVSLKAYAMNMPNSDELLKISIVAG 411
DB 195 EEDDEMDVBEERQBSGKRVAVRTAIN--EQOQQLKHYSILNRPNSDERMTAAARQ 252
QY 412 LPOEFVKEMF-----QRKVQYSNRS-----PSLETSKPLAENSPPTKD 454
DB 253 LDPRVVQVVFQNNRSERKMGQFQNNQAAGAAPMPIDSOASLTREDPL-----D 303
QY 455 SLLRSPVPMQSIPTSPSIAELHNSVTSCDPLRLTKSHFTNKAVD-KLHRSRNTPS 513
DB 304 LSVKBDPLTPKSESPPYIA-----PP-----SGEALNPEALINSRKFSTASWS 348
QY 514 PNLSTSSKNSSSYTPNSPSSSEELQAP-----IDLSPKO-----MRBP 556
DB 349 PASISPSAAALYFGAAPPSPNSQLDSTPPSGAFLPGLPYMLPMSLPMBALPGR-P 407
QY 557 KGIATKNTKATKATINL--DHSVSSSSSENSDEPLNLTFKKFSNNSNLDKSNPFVGM 615
DB 408 GGDFAA--NHALNNSIKLDPYRGTSLSPGGSE-----KRSWRDD----- 445
QY 616 NPFSAKPLVTLPLPPQSAFPAPATMPVQTSIGLARPYGLDQMSFLPMAYTPTGAATF 675
DB 446 -----SRISHDEFGAVLMP----- 462
QY 676 ADMQRRKYQKRGQFGODLLDGAQDYMSGLDMDTSDSCLSRKKIKTTESGYACDLCDK 735
DB 463 ---KPRGALETHGAHD-----PDLP-----YVCDQCK 489
QY 736 TFOKSSSLRHKYEHTGKPHOCQICKKAFKXHLIHSRLHSGEKPYQCDKCGKRFH 795
DB 490 AFAKQSLARHKYEHSGQRPYQICEPKAFKXKHLIHSRLHSGEKPYQCDKCLRFH 549
QY 796 SSGYSQHMNRYSYCK 811
DB 550 SSGYSQHMNRYSYCK 565

```

RESULT 11
 Q8CSP2 PRELIMINARY; PRT; 189 AA.

AC Q8C8P2;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Zinc finger homeobox 1a.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK077901; BAC37053.1;
SQ SEQUENCE 189 AA; 21400 MW; 52DEB7A580E9E978 CRC64;

Query Match 73%; Score 365; DB 11; Length 189;
Best Local Similarity 44.6%; Pred. No. 3.8e-15;
Matches 82; Conservative 27; Mismatches 47; Indels 28; Gaps 6;

QY 776 RLHSGEKPYQCDKCGKRFSSGSGSYSGQNMNHRYSYCKRAEEREAAREAREKGLGPTL 835
DB 2 RLHSGEKPYQCDKCGKRFSSGSGSYSGQNMNHRYSYCKRAEEREAAREAREKGLGPTL 56
QY 836 --LNNRAYLQSTPGQYDSEERESMP--DSESEKEHEKGE--GYGK-LRRDGD 886
DB 57 PEVLATVHVGARASPSQADSDERESELTREEDSESEKEHEKGE--GYGK-LRRDGD 116
QY 887 EEEEESE 932
DB 117 EEEEESE 176
QY 933 DHEE 936
DB 177 ESEQ 180
RESULT 12
Q8C827 PRELIMINARY; PRT; 914 AA.
AC Q8C827;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Zinc finger protein 62.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK048619; BAC33395.1;
SQ SEQUENCE 914 AA; 104811 MW; 196A50D47205E9B9 CRC64;

Query Match 6.9%; Score 343.5; DB 11; Length 914;
Best Local Similarity 21.5%; Pred. No. 6.7e-13;
Matches 196; Conservative 125; Mismatches 324; Indels 265; Gaps 41;
QY 6 AGNRKFKTECGKAFKRYHKLHKLRLHSGEKPYQCDKCGKRFSSGSGSYSHISKKGIC 65
DB 210 SGEKCKDECKGKSFNYSVLDHQRHITGKEKPYQCDKCGKAFRNS-----G 257

QY 66 LISVNGRRNNIKTGSSP---NSVSSSPTNSAITQLRNKLENGKPLMSEQTG----- 115
DB 258 L-----RVNHRKIHITGKEKPYQCDKCGKRFSSGSGSYSHISKKGICAFITCR 312
QY 116 -LLKITEPLDND--KYVLMATGFGSSGSPFNNGGICATSPLVGHPAQSPPMHLGYGM 172
DB 313 TLNHRKS--IHFGKPKYQCDKCGKRFSSGSGSYSHISKKGICAFITCR 351
QY 173 EAPLLGPTNMSNLSEVOKVLO-----IVDVTVRQKMDCKTEPISKLGYHMDPCSQP 227
DB 352 EKPY-----ECDECKAFRNSGGLVHNRITGKEKPYQCDKCGKAFITCR 396
QY 228 EEOGVTSPNIPVGLPVVSHNGATKSIIDYTLKRYNEAKACLOSLTDSRQOISNIKEK 287
DB 397 -----GLAV-----HKSII--HPGKAHEKDCGKSPFNS 424
QY 288 LRTLDVTDKMTBNHSISTPFSCQFCKESPFGPIPLHGERLYCKMNEIKAVLOPHE 347
DB 425 -----LLQHKTI--HTGERPVYCDVCGKTFRNSGGLKVRRL-----HT 462
QY 348 NIVPNKAGVFDNRALL--LSVLSSEKGLTSPINPK-----DHMSVLKAYAMNMP 398
DB 463 GEKPKCDVC--GKAYISRSSLKNHKGIMGEKPKYCKSCKSFYSALHQRHITRE 520
QY 399 ---NSDELKISIAVGLPQEFVKEMFEQR--KYVQ--YSNRSRPSLERTSKPLAPNSP 450
DB 521 KPFQDECKGA-----FRNSGLKXHKRHTGERLYCKMNEIKAVLOPHE 566
QY 451 TTQSLDLP-RSPVKPMS-----ITSPSIAELANSVTSQDPLR--LTKSSHFTNIKAVD 502
DB 567 INHKSVMHGEKPFKDCDECKAFITRYTLN--HKHILQEKPKYCKSCKSFYSALHQRHITRE 625
QY 503 KLDSRSNTSPPLNLSISSKSHSSSYTPNSFSEBELQAPRLDLSLKQREPRGIAT 562
DB 626 KRVHTR--KPFQDECKAFRNSGGL-----LKVHKRI----- 656
QY 563 KNTKATISINLHNSVSSSSSENDEPLNLTPIKEF--SNSNLNKSNNPFGMNPFAK 621
DB 657 -----HTGEKPYQCDKCGKAFITRYTLN--HKHILQEKPKYCKSCKSFYSALHQRHITRE 688
QY 622 PLYTLPPOSAPPATMPPVQTSIPGLRPYGLDQMSFLPHMATYPTGATPADMQOR 681
DB 689 TSYTCDECKGAFSSRTLISHKRVHGEKPKYQCDKCGKAFITRYTLN--HKHILQEKPKYCKSCKSFYSALHQRHITRE 738
QY 682 RYQKQGFQDLDLDAQDYMVGSLDMDTSDSCLSRKIKITKTESGMVACDLCDTFOSS 741
DB 739 RIHGEKPYQCDKCGKAFRNSGGL-----TVNKRHTGK--PYGDCDECKAFITRYTLN--HKHILQEKPKYCKSCKSFYSALHQRHITRE 788
QY 742 SLNRKYEHTGKRPPOCOICKKAFKHKHLLHSHRLHSGEKPYQCDKCGKRFSSGSGSYQ 801
DB 789 SLINHSYHNRKQRYNCE--CKSPYRVSVDHQRHITGKEKPYQCDKCGKAFITRYTLN--HKHILQEKPKYCKSCKSFYSALHQRHITRE 847
QY 802 HNNHRYSYCKRAEEREAAREAREKGLGPTLLNNAVYLOSTPGQYDSEERESMPR 861
DB 848 H-----KRIHTGEESLMMANMESHSGTFQKMI--YEG-----GNSLDGTRMQNP 890
QY 862 DGESEKHEK 871
DB 891 --VWEAEHAK 898

RESULT 13
Q35700 PRELIMINARY; PRT; 949 AA.
AC Q35700;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Eviddelta 105.
OS Evi1 OR Evi-1.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bartholomew C., Clark A.M., Kilbey A.;
 RT "Identification and characterization of novel murine Evi-1 proteins.";
 RL Submitted (Sep-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=8311086; Pubmed=2842066;
 RA Morishita K., Parker D.S., Mucenick M.L., Jenkins N.A., Copeland N.G.,
 RA Thle J.N.;
 RT "Retroviral activation of a novel gene encoding a zinc finger protein
 in IL-3-dependent myeloid leukemia cell lines."
 RL Cell 54:831-840(1998).
 DR EMBL; AJ001482; CA04777.1; -
 DR HSSP; P08047; 1SP2.
 DR MGD; MGI:95457; Evi1.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00096; zf-C2H2_10.
 DR Prodom; PD000003; Znf_C2H2_4.
 DR SMART; SM00355; Znf_C2H2_10.
 DR PROSITE; PS00280; IG_MHC; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 10.
 DR Metal-binding; Zinc; Zinc-finger.
 KW SEQUENCE 949 AA; 106536 MW; 50754E1A26A158BB CRC64;

Query Match 6.8%; Score 341.5; DB 11; Length 949;
 Best Local Similarity 19.8%; Pred. No. 9,4e-13;
 Matches 211; Conservative 133; Mismatches 354; Indels 369; Gaps 44;

OY 9 RKEKTECGKAFKFKHKLKEHLRIHSGEKPYECNCKGRFSHSGSYSHISKKICGLIS 68
 DB 101 REIKCQCQCPAFKFWKSLIHHOMSHDSGKIYECENCKAVTTPDSNLRHRSQHV----- 155
 OY 69 VNGRMNNIKTSGSSPNVSSSPTNSAITQLRNKLNGKPL-----SMSEGTGLIKIT 121
 DB 156 -----GAAHACPECGKTFATSSGKQKHHSVKPFICEVCHSKSYQFNSLCHKR 208
 OY 122 EPLD-----FNDYKVLMTATGFSG-----SSPPM----- 145
 DB 209 MHADCRTOIKDCQGMFTSTSLINKRRCEGKNHFAAGFFGQGISLPTGTPAMDKTSM 268
 OY 146 -----NGGE-----GATSP-----GVHPSAOSPMDHGLGVMEAPLIGF 179
 DB 269 VNNSHANPGLADYFGTKRHPAGLTFPPAPGFSFSFGLFPS-----GLYHRPPLIPA 320
 OY 180 PTNNSNLSEVQKVLQIVNTVSRQMDCKTEDISKLGYHMKDPCSQPEQGVTSFVNIIP 239
 DB 321 SPVVKGLSTEGSNKCGSPLLTHPQILPATQDILKALSKL-----PP 362
 OY 240 VG-----LPVYSHGATGSIIDYTEKNE-AKACLOSLLTDSRRQISNKKELRT 230
 DB 363 VGDNKPVLELP-----ERSSSERPLEKSIDSESSDLDVSTPSSGSDLETSSGSDLES 415
 OY 291 LILVLVTDMMKIEMHSISTPSCQCFKSPGPPIPLHGHERYLCKMNEIKAVLQPHENI- 349
 DB 416 --DLSDSKER-----CKEN-----GMFMDKSPDLNLA 442
 OY 350 -VENKGVFVNKALLLSVLSSEKGLTSPIN-PYKDHMSYLKAYVANNMPPNSDELKIS 407
 DB 443 SINNKKE--HNNSVFSASVEQSAVSGAVNDSIKALASIAEKYF-----SGTGLVGLQ 494
 OY 408 I-----AVGLDQEFYKWEF--QRKVQYNSRSRSPSLERTSKPLAENSNTTKOSLLRSP 461
 DB 495 DKYVGLALPYPSMPLPFPPAFPSQSMYPF-----PDRDLRSILPL-----KMEQSP 539
 OY 462 VKRMDSTSPSIAMHNSVTSQCPPLRLTKSSHFTNKAVDKLDHSSNTPSPPLNLSSTS 521
 DB 540 -----SEVKKLQKSSSESPPDLITRKD 562

OY 522 SKNSHSSSTYPSFSSSELOAEPLDLSLPKQREPKGIATNKKTKATYSINDHNSVSS 581
 DB 563 EKPILSGPSKPSGTRATS-QDQPLDSMGSRG-----ASGTYLTERKXNHVGEKK 613
 OY 582 SENSND-EPNLNLFIFIKKEFSNNLNKSNNPVFGNMPF---SAPPLYTLPPQSAFPAT 637
 DB 614 GSNMIDRP-----SSDGLQHARTPPF-MPIYVAKRKLTDLLEALKE----- 657
 OY 638 FMPVYQTSIPGLRPYPGLDQMSFLPHMAYTPGATFADMOQRKQKQGFQDGLDG 697
 DB 658 -----KYLRFPSG-----FLTHPQPLP-----DQR----- 678
 OY 698 AADYNSGLDMDTDSBSCSLRKIKTE-----SCMYAD 731
 DB 679 --TWMSAINEMAEKESFALPPEASSELLQSVSPMSFPAPNTLPENILRKGRKERTYR 736
 OY 732 LCDKTFQKSSLLRHKYHTGKRPHQCOICKAFKHKHLIHSR-LHSGEKPYQCDKCG 790
 DB 737 YGKLFPRSANLRLHRTHTGEOYRCKYCDRSFSSIMQHVNNIHNKXPKCHLCD 796
 OY 791 KRFSSGYSQHMNHRYSYCKREAREAREAREKGH/GPTLLMNRAYLQSIIPQGY 850
 DB 797 RCGGQGTNIDRL-----KHHNGMSGT-----ATSSPH 826
 OY 851 SDSERESEMPRQGESEKEHEKE---GEQYGLRLRRDGESESESESESESESESESESE 905
 DB 827 SELESAGAILDDKEDAYFTEIRNFIGNSHNGSQSPRN-MEEMNGSGHPYDKKALATSONS 885
 OY 906 ETIRDEEFTGDHSDOSSSED-----GKME-----TSDHE--EDNMED 941
 DB 886 DLIDDEVEDEVALDDEDEDNDIPGKPRKELGVTRLDEIIPBDYHE 932

RESULT 14

OASD1 PRELIMINARY; PRT; 743 AA.

AC OASD1;
 DT 01-JUN-2002 (Tremblrel). 21, Created)
 DT 01-JUN-2002 (Tremblrel). 21, last sequence update)
 DT 01-MAR-2003 (Tremblrel). 23, last annotation update)
 DE Hypothetical 84.9 kDa protein.
 GN ZFP62.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RL EMBL; BC022935; AA022935.1; -
 DR MGD; MGI:99662; Zfp62.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00096; zf-C2H2; 22.
 DR SMART; SM00355; Znf_C2H2; 22.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 22.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 22.
 KW Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
 KW SEQUENCE 743 AA; 84942 MW; D1482B5602E5216 CRC64;

Query Match 6.8%; Score 331.5; DB 11; Length 743;
 Best Local Similarity 20.3%; Pred. No. 2,8e-12;
 Matches 166; Conservative 89; Mismatches 254; Indels 307; Gaps 29;

OY 6 AGNRKFKTECGKAFKFKHKLKEHLRIHSGEKPYECNCKGRFSHSGSYSHISKKICG 65
 DB 203 GSKKNCCKCEGSGSFYSSVLDOHKRIHGEKPYECGEGKAPRNS-----G 250
 OY 66 LLSVNGRMNNIKTSGSP---NSVSSPTNSAITQLRNKLNGKPLMSBQTS----- 115
 DB 251 L-----RVAKRIHTGKPYECOTCGTFPSNSGLRHHKRIHGEKPYECDECKAFTTCR 305
 OY 116 -ILKITEPLDFND--YKVLMTATGFSGSSPFMNGGLGATSPLGVAQSPMDHGLGVGM 172

```

Db      TLNHNKS--HFQGDPRYKCEDECEKSFNYSLLI-----QHKVHTG--- 344
Qy      173 EAPLLGPTMNSNLSVQKVLQ-----IYDNTVSQKMDCKTEDISKLGYNMKDPCSQP 227
Db      345 EEPY-----ECBECGKAFRNSGLIYHKRIHGEKRYKCDICGKAFSYS----- 389
Qy      228 EEOGVSPNIPVGLPVVSHNGATYSIIDYLEKVNKAACLOSLTDSRROIINIKKEK 287
Db      330 -----GLAV-----HKSI--HPGKAHECDCKSPSYNS----- 417
Qy      268 LRTLIDLVTDKMIENHSISTPSCQFCESFPGPPLPHQHERYLCNMKEIKAVLQPH 347
Db      418 -----LILQHKTI--HTGERPYVCDVCGKTRFNNSGLKVHRRLL-----HT 455
Qy      348 NIVPNKAGVFVDNKKALL--LSVLSSEKGLTSPINPYKDHMSVLKAYANMMENPDELKI 406
Db      456 GEKPYKCDVC--GKAYISHSLSKHKHGMGEKPYKCSYCEKSFNYSALB----- 504
Qy      407 SIAGVLPOEFYKEMFEQRKVQYNSRSPSLERTSKPLAPNSNPPTTKSLPRSPVKPMD 466
Db      505 -----QHKRI--HTREKPRGDCGCKAFRNSG----- 530
Qy      467 SITSPSIABLNSVTSQDPLRLTKSSHFTNIKAVDKLDHSRNTSPPLNLSSTSSKSH 526
Db      531 -----LKVHKKRIHTGERPYKCECG--KAY-----ISLSLHNKSV 565
Qy      527 SSSYTPNSFSSEELQAEPLDLSLPKOMREPKGIATKNTKATSIINLHNSVSSSENSD 586
Db      566 HGEKAP--FKCDECE-----KAFITRYTLNKKIHL-----GE 597
Qy      587 EPLNLTFLIKKEFSNNSNLDNKSNNPVFGMNPFSAPKELVTLPLPQSAFPAPATMPVQTSI 646
Db      598 KPYKCDVCEKSFN-----YTSLL----- 615
Qy      647 PGLRPYPLGIDMSFLPHMAVYTPGATPADMOQRKTKYORKGFGQDLIDGAQDYMSGLD 706
Db      616 -----SQRKRVHTREKPFEC-----RCE 634
Qy      707 DMTDSDCLSRKKIKKTESGMYACDLCDKTPQKSSSLRHKYEHGKPRHOCQICKKAFK 766
Db      635 KVFRRNSSLKHKRIHTEKPYECDICGKAYISHSLSLHNKSTHPGKTSYTCDECGKAF 694
Qy      767 HKHHLIHSRLHSGEKPYQCDKCGKRFSGSGYSQ 802
Db      695 SSRLLISHKRVHLGKPFKVCCEKSFYSLSLQ 730

RESULT 15
Q8TBA9 PRELIMINARY; PRT; 751 AA.
AC Q8TBA9;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Hypothetical protein.
OC Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC022992; AAH22992.1; -.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR001087; Znf_C2H2.
DR Pfam; PF01352; KRAB; 1.
DR Pfam; PF00096; zf_C2H2; 19.
DR ProDom; PD000003; Znf_C2H2; 18.
DR SMART; SM00349; KRAB_1.
DR SMART; SM00355; Znf_C2H2; 19.

```

```

DR PROSITE; PS08005; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 19.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 19.
DR Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 751 AA; 86190 MW; A05FCECBF1275CB CRC64;

Query Match
6.6%; Score 331.5; DB 4; Length 751;
Best Local Similarity 18.1%; Pred. No. 2.9e-12;
Matches 148; Conservative 77; Mismatches 198; Indels 393; Gaps 27;

Qy      7 GNRKFKTECGKAFYKHHKLEHLRIHSGEKPYECPNKKRPSHSGYSYSHSKKICGL 66
Db      302 GEKPYKCECGKAFSQRTHLVQHORIHTEKRYTNECGKAFSQRGHMEH----- 352
Qy      67 ISVNGRMNNIKTSSP--NSVSSPNSAITQLRNLTENGKPLMSBQGTGLKITEP 123
Db      353 -----QKIHTEKPYKCEBCDCTFTRSHLTQHKIHGE-----KT-- 389
Qy      124 LDFNDYKVLMTGFGSGSPMNGIGAT--SPLGVHPSAOSPMQHLGVMEAPLLGEP 181
Db      390 -----YKNECGKAFNPGSTFIRHMIHTEKPYECNECGKAFSQH----- 430
Qy      182 MNSNLSVQKVLQIYDNTVSRQKMDCKTEDISKLGK-----HKMDPCSQDEQVTS 235
Db      431 --SMLTQHK-----YTTGERPYDC--ABCGKSFYSYSLAQHKL----- 466
Qy      236 NIPVGLPVSHNGATKSIIDYLEKVNKAACLOSLTDSRROIINIKKEKLTLLIDLV 295
Db      467 -----IHTGK-----PYKNECGKAFSYCSLT--QHKRI----- 495
Qy      236 TDDKMIENHSISTPSCQFCESFPGPPLPHQHER-----YLCNMKEIKAVLQPHEN 348
Db      496 -----HTREKPREGCECGKAFSYLSNLQHQHTQHEKAYEYK----- 533
Qy      349 IVPKAGVFVDNKKALLSSVLS--KGLTSPINPYKDHMSVLKAYANMMENPDELKIS 407
Db      534 --ECG-----KAFIRSSSLAKHERIHTGERPYQCH----- 561
Qy      408 IAVGLPOEFYKEMFEQRKVQYNSRSPSLERTSKPLAPNSNPPTTKSLPRSPVKPMD 467
Db      562 -----ECGTFSTYG-----SLIQ----- 575
Qy      468 ITSPSIABLNSVTSQDPLRLTKSSHFTNIKAVDKLDHSRNT--PSPPLNLSSTSSKSH 526
Db      576 -----HKRIHTEKPYKNECGKAFN--QNIHLTQHKRIHTEKPYEACGCKAFRH 625
Qy      527 SSSYTPNSFSSEELQAEPLDLSLPKOMREPKGIATKNTKATSIINLHNSVSSSENSD 586
Db      626 CSS-----LAQHK-----HTEE 639
Qy      587 EPLNLTFLIKKEFSNNSNLDNKSNNPVFGMNPFSAPKELVTLPLPQSAFPAPATMPVQTSI 646
Db      640 KPYQCNKCEKTFSSSHLTQHOR-----IHT----- 665
Qy      647 PGLRPYPLGIDMSFLPHMAVYTPGATPADMOQRKTKYORKGFGQDLIDGAQDYMSGLD 706
Db      666 -GEK----- 669
Qy      707 DMTDSDCLSRKKIKKTESGMYACDLCDKTPQKSSSLRHKYEHGKPRHOCQICKKAFK 766
Db      670 -----YKNECGKAFSQRTHLVQHORIHTEKRYTNECGKAFSQRGHMEH----- 708
Qy      767 HKHHLIHSRLHSGEKPYQCDKCGKRFSGSGYSQ 802
Db      709 QSTYLIQHORLHSGEKPYGCDGKSFYRSALNKH 744

```

Search completed: January 5, 2004, 16:31:02
Job time : 54 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2004, 16:30:01 ; Search time 21 Seconds
(without alignments)
1901.974 Million cell updates/sec

Title: US-09-964-238-2

Perfect score: 4991
Sequence: 1 MLTQAGNRKFKTECGKAF.....DGKMETKSDHEEDNEDMGW 944

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/1/aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4991	100.0	944	US-09-449-285A-2	Sequence 2, Appli
2	276.5	5.5	706	US-08-074-967-2	Sequence 2, Appli
3	276.5	5.5	706	US-08-553-541B-2	Sequence 2, Appli
4	276.5	5.5	706	US-09-268-202-2	Sequence 2, Appli
5	276.5	5.5	706	PCT-US94-0669-2	Sequence 2, Appli
6	274	5.5	496	US-08-224-482-2	Sequence 2, Appli
7	274	5.5	533	US-08-040-548-1	Sequence 1, Appli
8	274	5.5	533	US-08-466-344-1	Sequence 1, Appli
9	274	5.5	533	5206152-2	Patent No. 5206152
10	272.5	5.5	543	US-08-224-482-4	Sequence 4, Appli
11	272.5	5.5	543	US-09-702-705-337	Sequence 337, App
12	272.5	5.5	543	US-09-736-457-337	Sequence 337, App
13	258	5.2	51	US-09-449-285A-21	Sequence 21, Appli
14	256.5	5.1	1706	US-08-459-568-2	Sequence 2, Appli
15	256.5	5.1	1706	US-08-399-411-2	Sequence 2, Appli
16	256.5	5.1	1706	US-08-516-859A-2	Sequence 2, Appli
17	256.5	5.1	1706	US-09-586-472-2	Sequence 2, Appli
18	256.5	5.1	1706	US-09-528-706-2	Sequence 2, Appli
19	248.5	5.0	1719	US-08-459-568-4	Sequence 4, Appli
20	248.5	5.0	1719	US-08-399-411-4	Sequence 4, Appli
21	248.5	5.0	1719	US-08-516-859A-4	Sequence 4, Appli
22	248.5	5.0	1719	US-09-586-472-4	Sequence 4, Appli
23	248.5	5.0	1719	US-09-528-706-4	Sequence 4, Appli
24	248	5.0	675	US-08-317-522A-9	Sequence 9, Appli
25	248	5.0	675	US-08-439-818A-9	Sequence 9, Appli
26	248	5.0	675	US-08-751-965-9	Sequence 9, Appli
27	248	5.0	675	US-08-738-975-9	Sequence 9, Appli

28	248	5.0	675	2	US-08-728-626-9	Sequence 9, Appli
29	248	5.0	675	3	US-08-808-599A-9	Sequence 9, Appli
30	235.5	4.7	387	2	US-08-224-482-8	Sequence 8, Appli
31	234.5	4.7	671	3	US-09-121-321-16	Sequence 16, Appli
32	234.5	4.7	671	3	US-08-933-803A-16	Sequence 17, Appli
33	230.5	4.6	488	2	US-08-933-750C-17	Sequence 17, Appli
34	230.5	4.6	488	3	US-09-234-613-17	Sequence 17, Appli
35	228.5	4.6	543	4	US-09-362-123A-4	Sequence 4, Appli
36	227.5	4.6	455	2	US-08-224-482-6	Sequence 6, Appli
37	227.5	4.6	456	1	US-08-040-548-2	Sequence 2, Appli
38	227.5	4.6	456	1	US-08-466-344-2	Sequence 2, Appli
39	226	4.5	728	2	US-08-475-844-5	Sequence 5, Appli
40	226	4.5	728	5	PCT-US95-08429-5	Sequence 5, Appli
41	221.5	4.4	727	2	US-08-475-844-9	Sequence 9, Appli
42	221.5	4.4	727	5	PCT-US95-08429-9	Sequence 9, Appli
43	220	4.4	470	2	US-08-946-241B-2	Sequence 2, Appli
44	220	4.4	470	3	US-09-309-053-2	Sequence 2, Appli
45	220	4.4	479	2	US-08-946-241B-9	Sequence 9, Appli

ALIGNMENTS

```
RESULT 1
US-09-449-285A-2
; Sequence 2, Application US/09449285A
; Patent No. 6313280
; GENERAL INFORMATION:
; APPLICANT: Vlaams Interuniversitair Instituut voor Biotechnologie
; TITLE OF INVENTION: SWAD-INTERACTING POLYPEPTIDES AND THEIR USE
; FILE REFERENCE: 2676-42320S
; CURRENT APPLICATION NUMBER: US/09/449,285A
; CURRENT FILING DATE: 1999-11-24
; PRIOR APPLICATION NUMBER: PCT/EP98/03193
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 97201645.5
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 944
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-449-285A-2
```

```
Query Match 100.0%; Score 4991; DB 4; Length 944;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 944; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLTQAGNRKFKTECGKAFKJHLKEHLRHSGEKPEPCNCKRFSHSGSYSHISS 60
DB 1 MLTQAGNRKFKTECGKAFKJHLKEHLRHSGEKPEPCNCKRFSHSGSYSHISS 60

QY 61 KKCIGLISVNGRMNNIKTSSPNSVSSPTNSAITQLRNKLENGKPLSMSEOTGLKIK 120
DB 61 KKCIGLISVNGRMNNIKTSSPNSVSSPTNSAITQLRNKLENGKPLSMSEOTGLKIK 120

QY 121 TEPLDFNFKYKLMARHGSSGSPFNNGLGATSPGVPSAQSPPMOHJGUGNEAPLLGFP 180
DB 121 TEPLDFNFKYKLMARHGSSGSPFNNGLGATSPGVPSAQSPPMOHJGUGNEAPLLGFP 180

QY 181 TMNSLSEVQKVLQVDTVTVRQKMDCTEDISKLKGHYMKDPCSQPEOGVTSNIPV 240
DB 181 TMNSLSEVQKVLQVDTVTVRQKMDCTEDISKLKGHYMKDPCSQPEOGVTSNIPV 240

QY 241 GLPVVSHNGATKSIIDYTLKRVNEAKCLQSLTDSRQISNIKKKRLTLLDVTDDKM 300
DB 241 GLPVVSHNGATKSIIDYTLKRVNEAKCLQSLTDSRQISNIKKKRLTLLDVTDDKM 300

QY 301 IENHSISPFSCQFCKESFPGIPLHQRHRYLCKRNEIKAVLQHEHIVPKAGVFDN 360
DB 301 IENHSISPFSCQFCKESFPGIPLHQRHRYLCKRNEIKAVLQHEHIVPKAGVFDN 360
```

```

QY 361 KALLSSVLSSEKGLSPINPYKDHMSVLAAYAMNPNSEDLKLSIAVGLFOEFVKEM 420
| | | | |
DB 361 KALLSSVLSSEKGLSPINPYKDHMSVLAAYAMNPNSEDLKLSIAVGLFOEFVKEM 420
QY 421 FQORRYOVSNSRSPLESTSKPLANSPTTQDLSLPPVPMDSINSPSIAELHNSV 480
| | | | |
DB 421 FQORRYOVSNSRSPLESTSKPLANSPTTQDLSLPPVPMDSINSPSIAELHNSV 480
QY 481 TSCDPLRLTKSSHFNINAKVDLHDSRSTSPPLNLSSTSSKSHSSSYTPNFSSEEL 540
| | | | |
DB 481 TSCDPLRLTKSSHFNINAKVDLHDSRSTSPPLNLSSTSSKSHSSSYTPNFSSEEL 540
QY 541 QAEPLDLSPKQMRPEKGIATNKTATISINDHNSVSSSSNSDPLNLTIFIKKFSN 600
| | | | |
DB 541 QAEPLDLSPKQMRPEKGIATNKTATISINDHNSVSSSSNSDPLNLTIFIKKFSN 600
QY 601 SNLNDKSNPNVGMPPSAKPLYPPLPOSAPFPATFMPVOTSIPLGAPYGLDOMSF 660
| | | | |
DB 601 SNLNDKSNPNVGMPPSAKPLYPPLPOSAPFPATFMPVOTSIPLGAPYGLDOMSF 660
QY 661 LPHMAYTPTGATFADMOORRYKQKQFQGLDLGADQYMSGLDDMTDSDCLSRKXI 720
| | | | |
DB 661 LPHMAYTPTGATFADMOORRYKQKQFQGLDLGADQYMSGLDDMTDSDCLSRKXI 720
QY 721 KTESGMVACDLCDKTPQKSSSLRHKYHTGKRPQCOICCKAFYKHHLIHSRLHSG 780
| | | | |
DB 721 KTESGMVACDLCDKTPQKSSSLRHKYHTGKRPQCOICCKAFYKHHLIHSRLHSG 780
QY 781 EKPYQCDKCKGRSHSGSYQHNNHRYSYCKRAEEREAERARERKGHGPELLMNA 840
| | | | |
DB 781 EKPYQCDKCKGRSHSGSYQHNNHRYSYCKRAEEREAERARERKGHGPELLMNA 840
QY 841 YLOSITPOGYSDEERESMPRDESEKEHEKEGEYGLKRRDCDEEESESEENKS 900
| | | | |
DB 841 YLOSITPOGYSDEERESMPRDESEKEHEKEGEYGLKRRDCDEEESESEENKS 900
QY 901 MOTDPTTIRDEEBTGSHMSDSESDGEMETKSDHEEDNMEDGNG 944
| | | | |
DB 901 MOTDPTTIRDEEBTGSHMSDSESDGEMETKSDHEEDNMEDGNG 944

```

```

RESULT 2
US-08-074-967-2
; Sequence 2, Application US/08074967
; Patent No. 5641672
; GENERAL INFORMATION:
; APPLICANT: Dalla-Favera, Riccardo
; TITLE OF INVENTION: CLONING AND USES OF THE GENETIC LOCUS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/074,967
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 26,678
; REFERENCE/DOCKET NUMBER: 0575/43771
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550

```

```

; TELEFAX: (212) 664-0525
; TELEEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 706 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-074-967-2

```

Query Match 5.5%; Score 276.5; DB 1; Length 706;

Best Local Similarity 21.1%; Pred. No. 2.1e-13; Matches 126; Conservative 87; Mismatches 244; Indels 141; Gaps 22;

```

QY 259 LKRVNKAQKQSLTTSRROISNIK-KEKRLTLDLVTDD-----KMLENH----- 304
| | | | |
DB 112 LQMEHVVDCKRFRKIKSEAMVSAIKPREEFNLNMLPQDIMAIVRGVEVNNPLRS 171
| | | | |
QY 305 -----SITPPSCQCKESFPCGPIPLHGH-ERYLCKMNEIKAVIQLPHE 347
| | | | |
DB 172 AGCGESRAAPSLVSGLSLP-----PASYSWYSHLPVSSLSLFSDEEFADVMPVA 221
| | | | |
QY 348 NIVPNKAGVFNKALLSSVSEKG---LSPINPYKDHMSVLAAYAMNPNSEDEL 403
| | | | |
DB 222 NPPPK-----RALPCDAPVPEGEYRPTLEVPVNVCHSVITYSPKETTPPEARSD-- 272
| | | | |
QY 404 LKSIIVGL-----PQEFVKEFQORRYOVSNSRSPLESTSKPLAPNSPTTKDLSL-P 458
| | | | |
DB 273 MHTVVAELGKPAFARNAAPFPDCAKSK-EERPPSEDEILHFPFPAAPLNRKGLVSP 331
| | | | |
QY 459 RSPVK-----PMDITSPIAELHNS-----VTSQDPLRLTKSSHNTINIKAVDKLDH 506
| | | | |
DB 332 QSPQKSDCCPNSTPEKSSKNACILQSGSPPAKSPTPDKACSWKKYKFIIVNLSNQ--N 389
| | | | |
QY 507 SRNTPSPILNLSSTSSKSHSSYTPNFSSEELQAEPLDLSPKQMRPEKGIATNKT 566
| | | | |
DB 390 AKGGEQAEQLGLSLR-----AYTAPACQPPMEENLDLQSPYL----- 431
| | | | |
QY 567 KATSIINDHNSVSSSESDPLNLTIFIKKFSNSNLDKSNPNVGMPPSAKPLYP 626
| | | | |
DB 432 -----SAGEDSTIP-----QASRLNINIRSMTG-----SPRSSSHSP 467
| | | | |
QY 627 LPPQSAFPATFW-PPVQTSIGLARPYGLDOMSFLPHMAYTPTGATFADMOORRYK 685
| | | | |
DB 468 L-----YHMPKCTSCSGQSP-----QHAEMCLHTAGPTFAE-----EMGE 503
| | | | |
QY 686 RKQFGQGLDLGADQYMSGLDDMTDSDCLSRKXIKKTESGMVACDLCDKTPQKSSSLR 745
| | | | |
DB 504 TQSEYSDSCENGAFPCNCDKRFSEASLKRHTTLQTHSDKPYKCDQASFRYKGNLAS 563
| | | | |
QY 746 HKYEHGKRPQCOICCKAFKHKHHLIHSRLHSGEPYQCDKCGRFSHSGSYQHM 803
| | | | |
DB 564 HKTIVTGERPKYCNICGAQPNRPAKHTHRIHSGEKPKYCETCGARFQVAILRAHV 621
| | | | |

```

```

RESULT 3
US-08-553-541B-2
; Sequence 2, Application US/08553541B
; Patent No. 5882858
; GENERAL INFORMATION:
; APPLICANT: Dalla-Favera, Riccardo
; TITLE OF INVENTION: CLONING AND USES OF THE GENETIC LOCUS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10036
; COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553,541B
FILING DATE: May 28, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/43771-A-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELETYPE: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 706 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-553-541B-2

Query Match 5.5%; Score 276.5; DB 2; Length 706;
Best Local Similarity 21.1%; Pred. No. 2.1e-13;
Matches 126; Conservative 87; Mismatches 244; Indels 141; Gaps 22;

259 LEKNEAKACILOSITTSRROISNIK--KEKLRITLIDVTD-----KMIENH----- 304
112 LQMEHVVDTCRKFIKASAEWWSAIKPREEFELNSRMLMPDIMA YRGREVENNLPLRS 171
305 -----SISTPPSCQCKESFPGPIPLHOH--ERYLCKNNEIKAVLOPHE 347
172 APGESRAFAFSLVSGLSTP-----PASYSMTSHLPVSSILFSDDEFRDVRMVA 221
348 NIVNKAAGVFDNKKALLSVLSBK-----LTSPIINPKDMSVLKAYANMEBNSDEL 403
222 NPPFE-----RALPCDSARPYGEXSRPTLEVSPTVCHSNISYSPKETIPEEARSD-- 272
404 LKISIAVGL-----POEFYKWEFQKRYQYNSRSPSLERTSKPLAPNSPTTQDLSL-P 458
273 MHSVABGLKPAAPFARNAPYFPCDKASK--BEERPSSEDEIALHPEPPNAPLNKGLVSP 331
459 RSPVK-----PMDSTSPSIABLHNS-----VTSCEPPLRLTKSSHFTNIKAVDKLDH 506
332 QSPKSDCQPNSTPEACSSKNACILQSGSPPAKSPDPKACSWKKYFIVLNSLNG--N 389
507 SRSNTPSPPLNLSSTSSKNSHSSYTPNSFSSSELOAEPLDLSLPKQMEPKGIATKNKT 566
390 AKPGGPEGAELGRLSR-----ATTAPACOPPEPEPNDLQSPTKL----- 431
567 KATSLINDHNSVSSSSSENSEDEPLNLTFTKKEFSNNLNKNSNPVPGMNPFSKAPLYTP 626
432 -----SASGEDSTIP-----QASRLNNIVNRSMTG-----SPRSSSESHSP 467
627 LPPGSAFPAPATFM--PPVQTSIPGLRPYGLDQMSFLPHMATYTPPGATPFADMOQRKYQ 685
468 L-----YMHPPKCTSCGSOSP-----QHAEMCLHTAGPTFAE-----EMGE 503
686 RKQGFQGLDLGAQDYMSGLDMDTSDSCLSRKKIKTESGMYACDLCDKTFQKSSSLR 745
504 TQSEYSDSSCENGAFCNMBDCRFSSEASIKRHTLQTHSDKRYKCDRQASFRYGNLAS 563
746 HKYHTETGRPHOCQICCKAFKHKHLLIHSRLHSEKRYQCDCKGRPSHSGSYQHM 803
564 HKTATHTGKPYKCNICGAFNRPALIKTHTRISHGEKRYKCTCGARFVQVAHLRAHV 621

RESULT 4
US-09-268-202-2
Sequence 2, Application US/09268202
Patent No. 6174997

GENERAL INFORMATION:
APPLICANT: Dalla-Favera, Riccardo
APPLICANT: Chaganti, Raju S.K.
TITLE OF INVENTION: CLONING AND USES OF THE GENETIC LOCUS
TITLE OF INVENTION: bcl-6
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/268,202
FILING DATE:
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/43771-A-PCT-US-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELETYPE: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 706 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-268-202-2

Query Match 5.5%; Score 276.5; DB 3; Length 706;
Best Local Similarity 21.1%; Pred. No. 2.1e-13;
Matches 126; Conservative 87; Mismatches 244; Indels 141; Gaps 22;

259 LEKNEAKACILOSITTSRROISNIK--KEKLRITLIDVTD-----KMIENH----- 304
112 LQMEHVVDTCRKFIKASAEWWSAIKPREEFELNSRMLMPDIMA YRGREVENNLPLRS 171
305 -----SISTPPSCQCKESFPGPIPLHOH--ERYLCKNNEIKAVLOPHE 347
172 APGESRAFAFSLVSGLSTP-----PASYSMTSHLPVSSILFSDDEFRDVRMVA 221
348 NIVNKAAGVFDNKKALLSVLSBK-----LTSPIINPKDMSVLKAYANMEBNSDEL 403
222 NPPFE-----RALPCDSARPYGEXSRPTLEVSPTVCHSNISYSPKETIPEEARSD-- 272
404 LKISIAVGL-----POEFYKWEFQKRYQYNSRSPSLERTSKPLAPNSPTTQDLSL-P 458
273 MHSVABGLKPAAPFARNAPYFPCDKASK--BEERPSSEDEIALHPEPPNAPLNKGLVSP 331
459 RSPVK-----PMDSTSPSIABLHNS-----VTSCEPPLRLTKSSHFTNIKAVDKLDH 506
332 QSPKSDCQPNSTPEACSSKNACILQSGSPPAKSPDPKACSWKKYFIVLNSLNG--N 389
507 SRSNTPSPPLNLSSTSSKNSHSSYTPNSFSSSELOAEPLDLSLPKQMEPKGIATKNKT 566
390 AKPGGPEGAELGRLSR-----ATTAPACOPPEPEPNDLQSPTKL----- 431
567 KATSLINDHNSVSSSSSENSEDEPLNLTFTKKEFSNNLNKNSNPVPGMNPFSKAPLYTP 626
432 -----SASGEDSTIP-----QASRLNNIVNRSMTG-----SPRSSSESHSP 467
627 LPPGSAFPAPATFM--PPVQTSIPGLRPYGLDQMSFLPHMATYTPPGATPFADMOQRKYQ 685
468 L-----YMHPPKCTSCGSOSP-----QHAEMCLHTAGPTFAE-----EMGE 503

QY 686 RKQFQGGDLLDGAQDYMSGLDMDTSDCLSRKKIKTESGMVACDLCDTKFQKSSSLR 745
DB 504 TQSEYSDSCENGAFPCNECDCCFSEBASLKHHTLTQTHSDKPYKCDRCQASFRYKGNLAS 563
QY 746 HKYHTGKRPKPHOCQICKKAFFKHKHLLIHSRLHSGKPYQCDKCGKRFHSGSYQHM 803
DB 564 HKTVHTGKRPYKCNICGAQFNRPNALKHTYRHSKGKPYKCTCGARFQVAHLRAHV 621

RESULT 5
PCT-US94-06669-2
Sequence 2, Application PC/TUS9406669
GENERAL INFORMATION:
APPLICANT: Dalla-Favera, Riccardo
APPLICANT: Chaganti, R.S.K.
TITLE OF INVENTION: CLONING AND USES OF THE GENETIC LOCUS
TITLE OF INVENTION: bcl-6
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10112

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/06669
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/074,967
FILING DATE: 09-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 43771-A-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 706 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-06669-2

Query Match 5.5%; Score 276.5; DB 5; Length 706;
Best Local Similarity 21.4%; Pred. No. 2.1e-13;
Matches 126; Conservative 87; Mismatches 244; Indels 141; Gaps 22;

QY 259 LEKVNKAQALQSLTDSRQISNIK--KEKATLIDLVDL-----KMIENH----- 304
DB 112 LQMEHVVDTRKFKIKASEAMVSAIKPREEFINSRMLQDLMAYRAGRVRNENPLRS 171
QY 305 -----SISTPFSOCFKESPFGPIPLHQH--ERYLCMNEIKAVLQPH 347
DB 172 APCGESRAFAPSLYSGISTP-----PASYSMYSHLPVSSLLPSDEFPDVMFVA 221
QY 348 NIVNKAQVFNKALLSSVLESG-----LTSPINDYKDMSVLXAYYAMNNEPNSDEL 403
DB 222 NPPPKR-----RALPCDSARVPVGEYSRPTLEVSPNVCHSNISPKETIPREASD-- 272
QY 404 LKISIVGL-----PQEVKEMFEORKKYYQYNSNGRSPLETSKPLAMNSPTTKDILL-P 458
DB 273 MHYSVAGLKPAPASARNADYFPQDKASK--EEERPSSEDEIALHFEPNAPLNRKGLVSP 331

QY 459 RSPVK-----PMDSITSPSIAELANS-----VTSCEPPLRLTKSSHFTNIKAVDKLDH 506
DB 332 QSPQSDCQPNPTEACSPKACILQSGSGPPAKSPTPDKACSMWKYKFTVLSNLNQ--N 389
QY 507 SRNSNPSPLNLSSTSSKNSHSSSTYPNPSSESELQAEPLDLSLPKOMEPKQIINTKNT 566
DB 390 AKPGGEQAEQLRLPR-----AYTAPACQPPMEPENLDLQSPKTL----- 431
QY 567 KATSLNDHNSVSSSENSEDEPLNTLFIKKEPSNNSNLDKSNPNVPGMPPSAKPLVTP 626
DB 432 -----SAGEDSTIP-----QASRLNNIVNRMTG-----SPRSSSESHP 467
QY 627 LPPQSAFPPTATM--PVOYTSIPGLRPPGLDQMSFLPHAYTYTPGATFADMOQRKXY 685
DB 468 L-----YMHPPKTSQSGSP-----QHMEMLHTAGPTFAE-----EMGE 503
QY 686 RKQFQGGDLLDGAQDYMSGLDMDTSDCLSRKKIKTESGMVACDLCDTKFQKSSSLR 745
DB 504 TQSEYSDSCENGAFPCNECDCCFSEBASLKHHTLTQTHSDKPYKCDRCQASFRYKGNLAS 563
QY 746 HKYHTGKRPKPHOCQICKKAFFKHKHLLIHSRLHSGKPYQCDKCGKRFHSGSYQHM 803
DB 564 HKTVHTGKRPYKCNICGAQFNRPNALKHTYRHSKGKPYKCTCGARFQVAHLRAHV 621

RESULT 6
US-08-224-482-2
Sequence 2, Application US/08224482
Patent No. 5837692
GENERAL INFORMATION:
APPLICANT: Mercola, Dan
APPLICANT: Adamson, Eileen D.
TITLE OF INVENTION: Inhibition of the Mitogenic Activity of
TITLE OF INVENTION: PDGF by Mammalian EGF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/224,482
FILING DATE: 07-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-ME 9913
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 496 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-224-482-2

Query Match 5.5%; Score 274; DB 2; Length 496;
Best Local Similarity 25.4%; Pred. No. 1.9e-13;
Matches 99; Conservative 47; Mismatches 143; Indels 100; Gaps 14;

QY 518 SSTGS-----KNSHSSYTPNSFSSER-----LQAEPL-DLSLPPKQMEPKQIINTK 565

Db 27 SSTSGGGGGGSGSGSAGFNPQGEPSQPYEHLTTESFSDIALNNE---KAMVETSY 82
Qy 566 TKATISINDHNSVSSSSSENDEPL-----NLTFIKKEFNSNNLDNKSNNPVFGMNFSAK 621
Db 83 SQTTRL-----PPIYTGAFSLLEPAPNSGNTLMPBELFSLVSGLVSMTPPTSSSS----- 133
Qy 622 PLVPLPPOSAPFPATFMPVQTSIPG-----LRPYGLDQMSFLPHNAVITYPT 670
Db 134 ---APSPAASSSSASQSPPLSCAVPSNDSPISAAFTPTPTMD--IFPEPQOAPFG 188
Qy 671 GAATFADMQQRKTKQKQGF-----OGDLIDGADY--MSGLDMDTSDSC- 714
Db 189 SAGTALQYPPPAYPATKGFQVPMIPYLPFQOOGDLSLGTTPDQKPFQGLNRTQQPSLT 248
Qy 715 -LSRKIKKTESGM-----YAC--DLCDK 735
Db 249 PLSTIKAFATQSGSDLKALNTTYSQSLIKPSRMKTYPNRSKTPPHERPYACPVESCDR 308
Qy 736 TFOKSSSLRKHKEHTGRPHQCOICKAFKHKHLIHSRLHSGEKPEYQCDKCGKRFSH 795
Db 309 RFSRSDLTTHIRIHTGQKPFQCRICMNFSRSDHLTHIRHTGKEKPFACDICKRKFAR 368
Qy 796 SGYSQHMNRYSYCKRAEBEREAEREA 824
Db 369 SDEKRRHTKIHILROKDKKADKSVASPA 397

RESULT 7

US-08-040-548-1
; Sequence 1, Application US/08040548
; Patent No. 5763209
; GENERAL INFORMATION:
; APPLICANT: Sukhatme, Vikas P.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO THE
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESS: Arnold, White & Durkee
; STREET: 321 No. 5763209th Clark Street, Suite 800
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/040,548
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coughlin, Daniel F.
; REGISTRATION NUMBER: 36,111
; REFERENCE/DOCKET NUMBER: arc067
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 744-0090
; TELEFAX: (312) 245-4961
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 533 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-040-548-1

Query Match 5.5%; Score 274; DB 1; Length 533;
Best Local Similarity 25.4%; Pred. No. 2,1e-13;
Matches 99; Conservative 47; Mismatches 143; Indels 100; Gaps 14;
Qy 518 SSTSS-----KNSHSSSYTPNPSFSSB---LQAEPL-DLSLPKQMEPKGIATKXK 565

Db 64 SSTSGGGGGGSGSGSAGFNPQGEPSQPYEHLTTESFSDIALNNE---KAMVETSY 119
Qy 566 TKATISINDHNSVSSSSSENDEPL-----NLTFIKKEFNSNNLDNKSNNPVFGMNFSAK 621
Db 120 SQTTRL-----PPIYTGAFSLLEPAPNSGNTLMPBELFSLVSGLVSMTPPTSSSS----- 170
Qy 622 PLVPLPPOSAPFPATFMPVQTSIPG-----LRPYGLDQMSFLPHNAVITYPT 670
Db 171 ---APSPAASSSSASQSPPLSCAVPSNDSPISAAFTPTPTMD--IFPEPQOAPFG 225
Qy 671 GAATFADMQQRKTKQKQGF-----OGDLIDGADY--MSGLDMDTSDSC- 714
Db 226 SAGTALQYPPPAYPATKGFQVPMIPYLPFQOOGDLSLGTTPDQKPFQGLNRTQQPSLT 285
Qy 715 -LSRKIKKTESGM-----YAC--DLCDK 735
Db 286 PLSTIKAFATQSGSDLKALNTTYSQSLIKPSRMKTYPNRSKTPPHERPYACPVESCDR 345
Qy 736 TFOKSSSLRKHKEHTGRPHQCOICKAFKHKHLIHSRLHSGEKPEYQCDKCGKRFSH 795
Db 346 RFSRSDLTTHIRIHTGQKPFQCRICMNFSRSDHLTHIRHTGKEKPFACDICKRKFAR 405
Qy 796 SGYSQHMNRYSYCKRAEBEREAEREA 824
Db 406 SDEKRRHTKIHILROKDKKADKSVASPA 434

RESULT 8

US-08-466-344-1
; Sequence 1, Application US/08466344
; Patent No. 5773583
; GENERAL INFORMATION:
; APPLICANT: Sukhatme, Vikas P.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO THE
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESS: Arnold, White & Durkee
; STREET: 321 No. 5773583th Clark Street, Suite 800
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,344
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/040,548
; FILING DATE: 31-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Coughlin, Daniel F.
; REGISTRATION NUMBER: 36,111
; REFERENCE/DOCKET NUMBER: arc067
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 744-0090
; TELEFAX: (312) 245-4961
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 533 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-466-344-1

Query Match 5.5%; Score 274; DB 1; Length 533;

	Beet Local Similarity	25.4%; Pred. No.2.le-13.	
	Matches	99; Conservative	47; Mismatches 143; Indels 100; Gaps 14.
OY	518	SSTSS-----KSHSSSYTPNFSSEE---LQAEPL-DISLPKOMREPKGIATYTK	565
		: : : : : : :	
Db	64	SSTSGGGGCGGSGNSGSAPNQGEPSQPYEHLTTSEFSDIALNNE----KAWETSYVP	119
OY	566	TKAHSINLDHNSVSSSSSENDEPL----NLTFIKKEPFSNNLDMKSNPNPVGMNPFSAK	621
Db	120	SOITRL----PITYTGRFSLEBPADNSNTLMPEPLFIVSGSLVAMTNPFSSSS----	170
OY	622	PLYPLPPOSAFPRTFMVOTSIIG-----LRYPGLDQMSFLPHMAYTYPT	670
Db	171	--APSPAASSSSSKSQSPPFLSCAVFNSDSPIFYCAALFFRPNTD--IIFPPQSQAIRPG	225
OY	671	GAATPADMOQRKYORKOGF-----QGDLIDGAODY--MSGDLMDTSDSC-	714
Db	226	SAGTALQYPRPAYRATKGFQVPMYRDVLLPQQGGDLSGTFTDQKPRGLEMRTOQPBLT	285
OY	715	-LSRKIKKTESGM-----YAC--DLC DK	735
Db	286	PLSTIKAFATOSGDLLKALNTTYSQLIKPSRMKRYPNRPSEKTRPHERHYACPVSRCR	345
OY	736	TFOGSSILRHRYENTGKPRHOCCCKAKFKYKHNLIEHSRLHSEKRYQCDCGCKRPSH	795
Db	346	RFSNSDELTRHIRHTGQKPFQCRICMTNFSSDHLLTHIRHTGEBKPACDICGRKFAR	405
OY	796	SGSYSOHNHRYSYCRAEEEREALEREA	824
Db	406	SDERGHTTKIHRLROKDOKTADKSUVASPAW	434

```

RESULT 9
5206152-2
; Patent No. 5206152
; APPLICANT: SUKATNE, VIKAS P.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF EARLY GROWTH
; REGULATORY PROTEIN GENES
; NUMBER OF SEQUENCES: 18
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/249,584
; FILING DATE: 26-SEP-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 179,587
; FILING DATE: 08-APR-1988
; SEQ ID NO:2:
; LENGTH: 533
5206152-2

```

Query Match	5.5%	Score 274	DB 6	Length 533
Best Local Similarity	25.4%	Pred. No. 2.1e-13		
Matches	99	Conservative 47	Mismatches 143	Indels 100
			Gaps 14	
QY	518	SSTSS-----KNSHSSSYTPNFSFSEE----	LOAEPL-DLSPKOMEPPGIIATKK	565
Db	64	SSTSGGGGGGGSNSGSSAFNPQGESEBPYHLTTES	SDIALNNE-----KAMVETSY	119
QY	566	TKATSYINLDHNSVSSSSSENSDEPL----	NLTPIKKEFSNSNNLDKSNNPVFGMPFSK	621
Db	120	SQTRTL-----PITYTGRFLEAPNMSGNTLMP	LEPLFSLVSGLSVNTNPTSSS-----	170
QY	622	PLYTPPLPQSAFPATPTFMPEVOTSI	IFG-----LRPYGLDONSFLPHMAITYPT	670
Db	171	--ASPFAASSSSSSASQSPPLSCAVPNSDS	SPISAAAPTFPPNPTD--IFPEPQSGAFPG	225
QY	671	GAATFADMDQRRKRYQRKQGF-----	QGDLLDGDADY--MSGLDMDTBSDSC	714
Db	226	SAGTALQYPPPAYPATKGFQVPMIPDYLF	PQOQGDLSGTGDPQKPFQGLERTQOPSLT	285
QY	715	LSRKIKIKTESGM-----	YAC--DLCDK	735
Db	286	PLSTIKAPATQSGSDLKALNTTYSQLIKPSR	MRKRYPNRPSKTPPHERPYACPVESCCR	345

Oy	736	TFFKSSLLRNHYKENYNGKPHQCOJCSKAFGKHILLNHSRLSHGKEPYOOCDCGRPSH	795
	:	:::::	:
Dd	346	RFSRSELRFRHRIRIHGQKPFQCRIOMNFSPSDHLTHIRTHGEKXPACDICGRKFAR	405
Oy	796	SGSYSOHHNRYSYCREEAREEAREEA	824
	:	:::::	:
Dd	406	SDERKNHTKIHLQDKDKADKVASPPAA	434

RESULT 10
 US-08-224-482-4
 Sequence 4, Application US/08224482
 Patent No. 5837692
 GENERAL INFORMATION:
 APPLICANT: Mercola, Dan
 APPLICANT: Adams, Sileen D.
 TITLE OF INVENTION: Inhibition of the Mitogenic Activity of
 TITLE OF INVENTION: PDGF by Mammalian EGR
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Campbell and Flores
 STREET: 4370 La Jolla Village Drive, Suite 700
 CITY: San Diego
 STATE: California
 COUNTRY: USA
 ZIP: 92122
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/224,482
 FILING DATE: 07-Apr-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-ME 9913
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 543 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-224-482-4

Query Match	Similarity	5.5%	Score	272.5	DB 2	Length	543
Beet	Local	Similarity	26.3%	Pred	No. 2.8e-13		
Matches	96	Conservative	47	Mismatches	133	Indels	89
						Gaps	14
Qy	516	NLSSTSS-----KNSHSSSYTNSFSSE-----LQAEPL-DLSLPKOMEHPKCI	559				
		: :	: :	: :	: :	: :	: :
Db	61	NSSSSSSGGGGGGGGSSSSSSSTFNQADTGEQPEHLEHTLESPPDISLNN-----KYL	116				
Qy	560	IATNKRKTATSIINLDHNSVSSSSSENSDEPL-----NLTFIKKEFSNNINDKSNNPVFGM	615				
		: :	: :	: :	: :	: :	: :
Db	117	VETSYPSQOTRL-----PPIYTGRLGRSLBPAPNSGNTLMPEPLFLSVLSGLVSMTNPPASS	172				
Qy	616	---NPFSAKPLTYLPPOSAPPATFMPVPVQTSIGLAPYGLDMSLPLPHAAIYPTGA	672				
Db	173	SAPSPAASASASASQSPPLSCAVPSNDSSPIYSAAPTF--PTPMTD--IPEPOSQAPGSA	229				
Qy	673	ATFAMQQRKKYQRRQGF-----QGDLIDGADY--MSGIDDMTDSBC--L	715				
Db	230	GTALQYPPPAYPAAKGQVPMIPDYLPEQQGDGLGTPDGKPGQGLSSRTQPSLTP	289				
Qy	716	SRKKIKKTESGM-----YAC--DLCDKTF	737				
		: :	: :	: :	: :	: :	: :
Db	230	STIKAFATQSGQDLKALNTSYQSLIPSRMRKTPPNRPSKTPPHERPACVYESCDKRF	349				

QY 738 QKSSSLRHKYEHTGKRPPOCOICKAFKHKHLLIEHSRLHSGEKPYOCDKGRFSG 797
Db 350 SRSEDLTHIRIHGQKPFQCRICMRNFSRSDHLLTHIRHTGKPFACDICKRKFASD 409
QY 798 SYSQH 802
Db 410 ERKRH 414

RESULT 11

US-09-702-705-337
; Sequence 337, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaifanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Pan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 337
; LENGTH: 543
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-702-705-337

Query Match 5.5%; Score 272.5; DB 4; Length 543;
Best Local Similarity 26.3%; Pred. No. 2.8e-13;
Matches 96; Conservative 47; Mismatches 133; Indels 89; Gaps 14;

QY 516 NLSTSS-----KNSHSSSTYTPNSFSSE---LQAEPL-DLSLPKQREPRGI 559
Db 61 NSSSSSGGGGGGGGSGNSSSSTFNPQADTGEQPYEHLTAESFPDISLNNE---KVL 116
QY 560 IATNKTKATSIINDHNSVSSSENDEPL---NLTFIKKEFSNNLJDKSNPNVPGM 615
Db 117 VETISPGQTRL---PPTTYTGRFSLPAPNSGTTLMPEPLFSLVSGLVSMTPPASS 172
QY 616 ---NPFSAKPLYTLPPOSAPPATFMPVOTSIPLGRPYGLDOMSFLPHMAYTYPTGA 672
Db 173 SAPSPASASASQSPPLSCAVPSNDSSPIYSAFTF-PTPNTD--IFPEQSQAFPSA 229
QY 673 ATPADMOORRYQKQGF-----QGLLDGAQDY--MSGIDMTDSDC--L 715
Db 230 GTALQYPPAYPAKAGFOVPMIPDYLPQOQGLGTDPQKPFQGLSRTQPSLPL 289
QY 716 SRKIKKTESGM-----YAC--DLCDKTF 737
Db 290 STTAPATQSSQDLKALNTSYQSLIKPSMRKYPNRPSTPPHERPYACPVESCDRRF 349
QY 738 QKSSSLRHKYEHTGKRPPOCOICKAFKHKHLLIEHSRLHSGEKPYOCDKGRFSG 797
Db 350 SRSEDLTHIRIHGQKPFQCRICMRNFSRSDHLLTHIRHTGKPFACDICKRKFASD 409
QY 798 SYSQH 802
Db 410 ERKRH 414

RESULT 12
US-09-736-457-337
; Sequence 337, Application US/09736457
; Patent No. 6509448

; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaifanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Pan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 337
; LENGTH: 543
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-736-457-337

Query Match 5.5%; Score 272.5; DB 4; Length 543;
Best Local Similarity 26.3%; Pred. No. 2.8e-13;
Matches 96; Conservative 47; Mismatches 133; Indels 89; Gaps 14;

QY 516 NLSTSS-----KNSHSSSTYTPNSFSSE---LQAEPL-DLSLPKQREPRGI 559
Db 61 NSSSSSGGGGGGGGSGNSSSSTFNPQADTGEQPYEHLTAESFPDISLNNE---KVL 116
QY 560 IATNKTKATSIINDHNSVSSSENDEPL---NLTFIKKEFSNNLJDKSNPNVPGM 615
Db 117 VETISPGQTRL---PPTTYTGRFSLPAPNSGTTLMPEPLFSLVSGLVSMTPPASS 172
QY 616 ---NPFSAKPLYTLPPOSAPPATFMPVOTSIPLGRPYGLDOMSFLPHMAYTYPTGA 672
Db 173 SAPSPASASASQSPPLSCAVPSNDSSPIYSAFTF-PTPNTD--IFPEQSQAFPSA 229
QY 673 ATPADMOORRYQKQGF-----QGLLDGAQDY--MSGIDMTDSDC--L 715
Db 230 GTALQYPPAYPAKAGFOVPMIPDYLPQOQGLGTDPQKPFQGLSRTQPSLPL 289
QY 716 SRKIKKTESGM-----YAC--DLCDKTF 737
Db 290 STTAPATQSSQDLKALNTSYQSLIKPSMRKYPNRPSTPPHERPYACPVESCDRRF 349
QY 738 QKSSSLRHKYEHTGKRPPOCOICKAFKHKHLLIEHSRLHSGEKPYOCDKGRFSG 797
Db 350 SRSEDLTHIRIHGQKPFQCRICMRNFSRSDHLLTHIRHTGKPFACDICKRKFASD 409
QY 798 SYSQH 802
Db 410 ERKRH 414

RESULT 13
US-09-449-285A-21
; Sequence 21, Application US/09449285A
; Patent No. 6313280
; GENERAL INFORMATION:
; APPLICANT: Vlaams Internationaal Instituut voor Biotechnologie
; TITLE OF INVENTION: SMAD-INTERACTING POLYPEPTIDES AND THEIR USE
; FILE REFERENCE: 2676-4232US
; CURRENT APPLICATION NUMBER: US/09/449,285A
; CURRENT FILING DATE: 1999-11-24
; PRIOR APPLICATION NUMBER: PCT/EP98/03193
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 97201645.5
; PRIOR FILING DATE: 1997-06-02
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0

SEQ ID NO 21
 LENGTH: 51
 TYPE: PRT
 ORGANISM: SBD mutant
 US-09-449-285A-21

Query Match 5.2%; Score 258; DB 4; Length 51;
 Best Local Similarity 100.0%; Pred. No. 1e-13;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

166 OHGVGEAPLIGFPPTMNSNLSEVQKVLQIVDNTVSRQKMDCKTEDISKLG 216
 1 OHGVGEAPLIGFPPTMNSNLSEVQKVLQIVDNTVSRQKMDCKTEDISKLG 51

RESULT 14
 US-08-459-568-2

Sequence 2, Application US/08459568

Patent No. 5811304

GENERAL INFORMATION:

APPLICANT: Huang, Shi

TITLE OF INVENTION: Retinoblastoma Protein - Interacting

TITLE OF INVENTION: Zinc Finger Proteins

NUMBER OF SEQUENCES: 93

CORRESPONDENCE ADDRESS:

ADDRESSEE: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego

STATE: California

COUNTRY: USA

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/459,568

FILING DATE: 02-JUN-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/399,411

FILING DATE: 06-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-LJ 1264

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001

TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1706 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULAR TYPE: protein

US-08-459-568-2

Query Match 5.1%; Score 256.5; DB 2; Length 1706;
 Best Local Similarity 18.8%; Pred. No. 3.2e-11;
 Matches 200; Conservative 132; Mismatches 381; Indels 353; Gaps 41;

11 FKTECGAKFYKMHLEKRLHSG--EKRYECPNCKKRF-----HSGSYSHISK 61
 355 FPCQHCERKFAFGQERHMHISTINNAFKCKYCKGKFGQINRRHRRHETGLKRR 414
 62 KCIGLISVNGRKNRIKTKSSSPNSVSSPNSAITOLRNKLENGK-----LSMSEQT 114
 415 PSMTLQS-----SEDDDGKGVNVTSKDESSPQLGQDCILINSEKTI 456
 115 GLKIKTEPLDNDYKVLMTATGFGSSSPFM--NGGLGATSP-----LGYPHSAQSM 165
 457 -----SQEVL-----NSSFVEENGKELHPCCKYKAVPGTHINRRHQ 495

166 ----OHL---GVGEAPLIGFPPTMNSNLSEVQKVLQIVDNTVSRQKMDCKTEDISKLG 217
 496 RRVHERHLIPKGVRRKGLIEEPQPAEQAPPSQVYV-----
 218 YHMKPCSQPEEQGYT-----SPNIPVGLPVVSHNGATKS----- 253
 534 ----PSTPEEGETDDVYIMDISNISENLNYIIDGKIQTNSSTNSCNDVIEMESNAH 588
 254 ----IIDYLEKVENAKACLOSITTDSSR---QISNIKEKRLTILIDLVTDDKMIEN 303
 589 IYIGIDCLLPVVEITONIKSTOVAVTDLLDSDSSTNCESKRTTAPPLPKITET 648
 304 HSIPTFSCQCKESFPFGPIPLHQ---HERYLCKNNEIKAVLOPHENI----- 349
 649 ESDSTAPSCSL---SLPLSISTAEVVSFFKEKGVLYSSKLOLQOTDCLTLPAGFSAE 705
 350 VNNKAGVFDNKAALL---SVLSEKGLSPINPKDHVSUKAYAMMEEDSDLLAI 406
 706 IPLGPGVACASAPASMLPVTSRFRKRTSSPESP--QHSPLALRF----- 748
 407 SIAGVLPQEFVKEFEQRKYQYSNRSRPSLERTSKPLAPNSPTTKDLSLPRSPVPM 466
 749 ----GKPNCKXAMTD---IVLTSKPKLB--SRSDSPAMSLSGRDERETGSP----- 792
 467 STSPSIAELHNSVTSCDPLRLTK---SSHFTNIKAVDKLDHSRNTPTSLNLSSTSS 522
 793 ----PCFDEYKISKEMAASSTFSSV-----CNOQLDLSGK 826
 523 KNSHSSYTPNSFSSRELOAEPLDLSPKOMEKPIGIATKN-----KTKATSNL- 573
 827 QKSEGTGKTPVWES---VLDSVHKPKCDSSGKEKFNHLAQPAAKKKPTTCMLQ 880
 574 ----DHNSVSSSENDEPLNLTEIKKFSNNLNDK-----SNNPVF 613
 881 KYLMEYNKVSJLPTETTP-----VTSRSPCKSPDTPQDELGPDSGCVPTASPE 934
 614 GNNPFS-----AKPLYTPLPOS-----APPATMPPVOTSIPLG-- 649
 935 VVGPSSPPLQIASLSSGQLPLLTPTPESSPPCPVLTVAATPPPLLTPLVLSHSSDA 994
 650 ----RPYRGIDQMSFLPHMAVTP-----TGAATPADMQORRRYQRRQGP- 690
 995 SPQCCSPSPSNTAOSPLILSPVSPSPPIPVPEPLMASAPGPPTLSSSSSSSSSP 1054
 691 ----QGDLLDGA-----QDYSGLDMDTDSCTSRKK 719
 1055 SSSCSSTSPRPLAVSVSVSGDNLEASLPAVTFKQESSESGIKPKKEAPRAGGQSV 1114
 720 IKTBEGMYACDLCKTFQKSSSLRHKYEHTGKRPHQCICKAPKXKHLIHS-RLH 778
 1115 VOETFSKNFICNVCCSPFLSIDLTGKLSVHAEEWPFKCFVCQLFKVTDLSEHFLH 1174
 779 SEBKPYQCDKCGKRFSHSGSYSOHNNHRYSKREAEAREARERKGGHGLPELLMN 838
 1175 GVGNIFVCSVCKKEPAFLCNLQOH-----QRDLHPDEVCTHHEFSGTLRP----- 1220
 839 RAYLOSITPQGYSDSEERESMPRDG-ESEKEHE--KEGEGYGKLR 881
 1221 ----QNFIDPSKANYHMSLPEFPLTSRREBLNDSSELYTTIK 1262

RESULT 15
 US-08-399-411-2
 Sequence 2, Application US/08399411
 Patent No. 5811008
 GENERAL INFORMATION:
 APPLICANT: Huang, Shi
 TITLE OF INVENTION: Retinoblastoma Protein - Interacting
 TITLE OF INVENTION: Zinc Finger Proteins
 NUMBER OF SEQUENCES: 93
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/399,411
FILING DATE: 06-MAR-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1264
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1706 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-399-411-2

Query Match 5.1%; Score 256.5; DB 2; Length 1706;
Best Local Similarity 18.8%; Pred. No. 3,2e-11;
Matches 200; Conservative 132; Mismatches 381; Indels 353; Gaps 41;

QY 11 FKCTCGAFKFKYKHLKHLRHSQ--EKPYECPNCKKRF-----HSGSYSHISK 61
DB 355 PFCQCEKFKATKQGLERHMHITINAFKCKYCGKRFQINRRRHERHETGLKR 414
QY 62 KCIGLISVNGWRNNIKTGSSPNSVSSPTNSAITOLRNKLENGK-----LSMSEQT 114
DB 415 PSMTLQS-----SEDDDGKGNVTSKDESSPQLODCLINSEKT 456
QY 115 GLKTKTPELPNDYKVLMTATGFGSGSPFM--NGGLGATSP-----LGVHPSAQSPM 165
DB 457 -----SOEVL-----NSFVEENGVEKELHPCKYCKVFGTHIMRRHQ 495
QY 166 -----QHL-----GVGMEAPLLGFPPTMNSLSEVQXVLQIVDNTVSRQKMDCKTEDISKLG 217
DB 496 RRVHERHLIPKGVRRKGGLEBPAPPAQAPPSQNVYV----- 533
QY 218 YHMKDPCSQPEQGV-----SPNIPVGLPVSHNGATKS----- 253
DB 534 -----PSTEPEEGETDVIYIMDISNISENLNYIDGICQNSSTNSCDVIEMSNSAH 588
QY 254 -----IIVTYLKEVNEAKACLOSLTDSRR--QISNIKKEKRLTLIDVTDKMIEN 303
DB 589 LYGIDCLLTPTVVEITONIKSTOVSVTDLLKDSSTSCSKRRITASPPVLPIKTIET 648
QY 304 HSIPTFCQPKESFPGPIPLHQ-----HERYLCQNEEIKAVIQPHENI----- 349
DB 649 ESDSTAPSCSL--SLPLISTAEVVSFHKEGVYLSSKQLQLODCLTLPAGPSAAE 705
QY 350 VPKAGVVDNKAALL-----SSVLSSEKGLTSPINPYKDHMSVLKAYVAMNMBNSDELKI 406
DB 706 IPKLGVCASAPASMLPTVSSRFKRTSSPSSP--QHSPLALDF----- 748
QY 407 SIANGLPQEFVKEWFOQRKVQYNSRSRPSLERTSKPLAPNSPTTKDLSLRSPVKPM 466
DB 749 -----GKPNDKAAWTD-----TVLTSKKPKLE--SRSDSPAWSLSGRDERETGSP----- 792
QY 467 SITSPSIEIHNVSVCDEPLRLTK-----SSHFTNIKAVDKLDHRSNTPSPILNLSSTSS 522
DB 793 -----PCFDEYKISKEMAASTFSSV-----CNOQPLDLSGKV 826

QY 523 KNSHSSSTPNSFSSEELQAEPLDLSLPKQKREPKGIATQX-----XTKATSNIL- 573
DB 827 QKSBGTGKTPVPMWES-----VLDSVHKKPCDSEKGEFKNHILAQPAKKKKPTTCMLQ 880
QY 574 -----DHNSVSSSENDEPLNLTPIKKEFSNNLQK-----SNNVF 613
DB 881 KYLLNEIVGSLPTETPE-----VTRSPCKSPDTOPPELGDSSCSVPTAESPE 934
QY 614 GNNPFS-----AKPLYPLPPOS-----APPATFMPVQTSIPGL-- 649
DB 935 VVGSPSPPLQNASLSSGQLPPLITPTESSSPPCCPVLTVATPPPLPTVPLSHPSDDA 994
QY 650 -----RPYGLDQMSFLPHAYTP-----TGATPADWQQRKRYQKQGF- 690
DB 995 SPQCPSPFSNTTAQSPPLILSPVSPSPPIPVPEPLMSAASPPTLSSSSSSSSSFP 1054
QY 691 -----QGDLDGA-----QPYMSGLDMTSDSLSKK 719
DB 1055 SSSGSGTSPSPPLSAVSSVSSGDNLEASLPATVFKQEESESEGLKPKKEAPPAAGQSV 1114
QY 720 IKTESGMYACDLCDKTFQKSSSLLRHKYERTGKRP HQCQI CKAFKHKHLIEHS--RLH 778
DB 1115 VQETFSKNFICNVCESPPLSLIKDLTKHLSVAEEMFECFCVQLFKYKTDLSERFLH 1174
QY 779 SGEKPYQCDKCGKRPSSHSGSYSQHMNRYSYCKEAEEREAERERAREKHLGPTTELLMN 838
DB 1175 GVGNIFVCSVCKKEFAFLCNLQOH-----QRDLHPDEVCTHHFEFSGTLRP----- 1220
QY 839 RAYLOSITPQYSDESEKESMPROG--ESEKEHE--KEESEGYGKLR 881
DB 1221 -----QNTFDPKAVNEHMPSLPEEPLETREBEELNDSSEELYTTIK 1262

Search completed: January 5, 2004, 16:36:09
Job time : 25 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2004, 16:31:07 ; Search time 38 Seconds
(without alignments)
4984.350 Million cell updates/sec

Title: US-09-964-238-2
Perfect score: 4991
Sequence: 1 MLTQAGNRKFKCTCGKAF.....DGKMTKSDHEEDNWDG 944

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 733937 seqs, 200641211 residues

Total number of hits satisfying chosen parameters: 733937

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Published Applications AA:*
2: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4991	100.0	944	9 US-09-964-238-2	Sequence 2, Appli
2	2612	52.3	533	9 US-09-864-761-37151	Sequence 37151, A
3	2612	52.3	533	12 US-10-029-386-33350	Sequence 33350, A
4	1657.5	33.2	1125	10 US-09-974-298-114	Sequence 114, App
5	1647.5	33.0	1104	12 US-10-104-047-2506	Sequence 2506, Ap
6	1606.5	32.2	1154	15 US-10-233-561-10	Sequence 10, Appl
7	1566	31.4	1117	15 US-10-233-561-12	Sequence 12, Appl
8	1505	20.1	190	9 US-09-864-761-36988	Sequence 34882, A
9	1505	20.1	190	9 US-09-864-761-36988	Sequence 36988, A
10	356.5	7.1	1042	15 US-10-097-340-83	Sequence 83, Appl
11	355.5	7.0	1051	15 US-10-097-340-79	Sequence 79, Appl
12	349.5	7.0	1204	12 US-10-029-386-33060	Sequence 33060, A
13	329.5	6.6	269	9 US-09-864-761-38036	Sequence 38036, A
14	329.5	6.6	269	12 US-10-029-386-33074	Sequence 33074, A
15	318	6.4	1357	12 US-10-295-027-1199	Sequence 1199, Ap

16	299.5	6.0	924	15 US-10-007-280A-231	Sequence 231, App
17	297	6.0	610	10 US-10-094-749-2604	Sequence 2604, Ap
18	296.5	5.9	610	12 US-09-989-920-212	Sequence 212, App
19	294	5.9	648	12 US-10-094-749-2472	Sequence 2472, Ap
20	290	5.8	514	12 US-10-029-386-34112	Sequence 34112, A
21	282.5	5.7	576	12 US-10-094-749-2365	Sequence 2365, Ap
22	279	5.6	902	12 US-10-029-386-32110	Sequence 32110, A
23	278.5	5.6	742	12 US-10-029-386-31985	Sequence 31985, A
24	276.5	5.5	706	9 US-09-107-058-2	Sequence 2, Appli
25	276.5	5.5	706	9 US-09-761-117-2	Sequence 2, Appli
26	276	5.5	816	12 US-10-029-386-33143	Sequence 33143, Ap
27	275	5.5	545	12 US-10-144-194A-26	Sequence 26, Appl
28	275	5.5	811	12 US-10-144-194A-26	Sequence 98, Appl
29	275	5.5	811	12 US-10-144-194A-26	Sequence 36148, A
30	273.5	5.5	536	9 US-09-864-761-36148	Sequence 36148, A
31	273.5	5.5	727	15 US-10-097-340-81	Sequence 81, Appl
32	272.5	5.5	543	10 US-09-726-457-337	Sequence 337, App
33	272.5	5.5	543	10 US-09-902-941-337	Sequence 337, App
34	272.5	5.5	543	10 US-09-849-626-337	Sequence 337, App
35	272.5	5.5	543	11 US-09-476-300-337	Sequence 337, App
36	272.5	5.5	543	12 US-10-251-661-6	Sequence 337, App
37	272.5	5.5	543	12 US-10-113-872-337	Sequence 337, App
38	272.5	5.5	543	12 US-10-116-275-180	Sequence 180, App
39	272.5	5.5	543	12 US-10-341-434-184	Sequence 184, App
40	272.5	5.5	543	12 US-10-341-434-198	Sequence 198, App
41	272.5	5.5	543	15 US-10-017-754-337	Sequence 337, App
42	272.5	5.5	543	15 US-10-205-823-95	Sequence 95, Appl
43	270.5	5.4	614	9 US-09-864-761-43107	Sequence 43107, A
44	269.5	5.4	566	12 US-10-322-774-10	Sequence 10, Appl
45	269.5	5.4	718	15 US-10-097-340-77	Sequence 77, Appl

ALIGNMENTS

RESULT 1
US-09-964-238-2
; Sequence 2, Application US/09964238
; Patent No. US20020035246A1
; GENERAL INFORMATION:
; APPLICANT: Vlaams Interuniversitair Instituut voor Biotechnologie
; TITLE OF INVENTION: SWAD-INTERACTING POLYPEPTIDES AND THEIR USE
; FILE REFERENCE: 2676-4232US
; CURRENT APPLICATION NUMBER: US/09/964, 238
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 09/449, 285
; PRIOR FILING DATE: 1999-11-24
; PRIOR APPLICATION NUMBER: PCT/EP98/03193
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 97201645.5
; PRIOR FILING DATE: 1997-06-02
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 944
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-964-238-2
Query Match 100.0%; Score 4991; DB 9; Length 944;
Best Local Similarity 100.0%; Pred. No. 3.38-284;
Matches 944; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 MLTQAGNRKFKCTCGKAFKKKHLKHLRHSSEKYECCNCKKRSSHSSSHSS 60
1 MLTQAGNRKFKCTCGKAFKKKHLKHLRHSSEKYECCNCKKRSSHSSSHSS 60
1 KKCIGILVNGRMNRNITGSSPNVSSPTNSATITOLRNKLENGKPLSMSEOTGLKIK 120
1 KKCIGILVNGRMNRNITGSSPNVSSPTNSATITOLRNKLENGKPLSMSEOTGLKIK 120
1 TEPIDFNDYKVLMTATGFGSSSPFNGGLGATSPGLVHPASQSPMOHLGVGNEADLP 180

```

Db 121 TEPLDNDYKVLMTATGFGSSGSPFMNGGLATSP LGVHPSAQS PMOHLGVGMEAPLLGFP 180
Qy 181 TMSNLSSEVQKVLQIYDNTVSRQKMDCKTEDI SKLKGYNHMDCSOPEEGVTSPIPIPV 240
Db 181 TMSNLSSEVQKVLQIYDNTVSRQKMDCKTEDI SKLKGYNHMDCSOPEEGVTSPIPIPV 240
Qy 241 GLPVSHNGATKSIIDYTLERKNEAKACLOSLTTSRQISNIKKEKRLTIDLVTDDKA 300
Db 241 GLPVSHNGATKSIIDYTLERKNEAKACLOSLTTSRQISNIKKEKRLTIDLVTDDKA 300
Qy 301 IENHSTSPSCCFCKESFPGPILPHOHERYLCKMNEIKAVIOPHNITPNKAGVVDN 360
Db 301 IENHSTSPSCCFCKESFPGPILPHOHERYLCKMNEIKAVIOPHNITPNKAGVVDN 360
Qy 361 KALLSSVLSSEKGLTSPINPYKDHMSYLKAYAMNENPNDLKI SIAGVLPOEFYKEM 420
Db 361 KALLSSVLSSEKGLTSPINPYKDHMSYLKAYAMNENPNDLKI SIAGVLPOEFYKEM 420
Qy 421 FEORATYQVSNSSPSLERTSKPLAPNSPTTQSLIPSPVPKMDSTSPSIAELHNSV 480
Db 421 FEORATYQVSNSSPSLERTSKPLAPNSPTTQSLIPSPVPKMDSTSPSIAELHNSV 480
Qy 481 TSCDPLRLTKSHFTNFKAVDKLDHSRNTSP LINSSTSSKNSHSSSTPNSFSEEL 540
Db 481 TSCDPLRLTKSHFTNFKAVDKLDHSRNTSP LINSSTSSKNSHSSSTPNSFSEEL 540
Qy 541 QAEPLDLSLPKQREPKGIATKTKTATSI INLDHNSVSSSSSENDEPLNLTPIKKEFSN 600
Db 541 QAEPLDLSLPKQREPKGIATKTKTATSI INLDHNSVSSSSSENDEPLNLTPIKKEFSN 600
Qy 601 SNNLDKSNNPVGMNPFSAKPLYTPLP POSAFPPATMPVPVOTSTIPGLRPYGLDMQSF 660
Db 601 SNNLDKSNNPVGMNPFSAKPLYTPLP POSAFPPATMPVPVOTSTIPGLRPYGLDMQSF 660
Qy 661 LPHMATYPTGATTPADMOQRKRYKORQGFQGLDLSAOLYSGLDMDTSDSCLSKKI 720
Db 661 LPHMATYPTGATTPADMOQRKRYKORQGFQGLDLSAOLYSGLDMDTSDSCLSKKI 720
Qy 721 KKTESGMYACDLCTFKQSSSLRHNYEHTGKRPHOCQICKA FKIKHILHSHRLHSG 780
Db 721 KKTESGMYACDLCTFKQSSSLRHNYEHTGKRPHOCQICKA FKIKHILHSHRLHSG 780
Qy 781 EKRYQCDKCKRSHSGSYQHMNHRSYCKREAEEREAERAREKGHIGPTELLMNA 840
Db 781 EKRYQCDKCKRSHSGSYQHMNHRSYCKREAEEREAERAREKGHIGPTELLMNA 840
Qy 841 YLOSITPGQYSDEBERESMPRDSSEKHEKEGEGYGLRRRDGDEEESESEENKS 900
Db 841 YLOSITPGQYSDEBERESMPRDSSEKHEKEGEGYGLRRRDGDEEESESEENKS 900
Qy 901 MDTDPETIRDEETGSHMDSSSEDGKMETKSDHEEDNMDGNG 944
Db 901 MDTDPETIRDEETGSHMDSSSEDGKMETKSDHEEDNMDGNG 944

```

RESULT 2
US-09-864-761-37151
Sequence 37151, Application US/09864761
Patent No. US20020048763A1

GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wenheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26

```

; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37151
; LENGTH: 533
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010130.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
; OTHER INFORMATION: EST HUMAN HIT: A1879705.1, EVALUE 3.00e-87
; OTHER INFORMATION: SWISSPROT HIT: P37275, EVALUE 2.00e-45
US-09-864-761-37151

```

Query Match 52.3%; Score 2612; DB 9; Length 533;
Best Local Similarity 94.2%; Pred. No. 4e-145;

Matches 502; Conservative 12; Mismatches 19; Indels 0; Gaps 0;

```

Qy 36 EKRYECPNCKKRRSHSGSYSHSSKCTGILSYNGRMNNITKGS PMSVSSSPNNSAI 95
Db 1 EKRYECPNCKKRRSHSGSYSHSSKCTGILSYNGRMNNITKGS PMSVSSSPNNSAI 60
Qy 96 TOLRNLKNGKPLSMSEOTGLLKITKTEPLDFNDYKVLMTATGFGSSGSPFMNGGLGATSP 155
Db 61 TOLRNLKNGKPLSMSEOTGLLKITKTEPLDFNDYKVLMTATGFGSSGSPFMNGGLGATSP 120
Qy 156 GVHPSAQS PMOHLGVGMEAPLLGFP TMSNLSSEVQKVLQIYDNTVSRQKMDCKTEDI SKL 215
Db 121 GVHPSAQS PMOHLGVGMEAPLLGFP TMSNLSSEVQKVLQIYDNTVSRQKMDCKA EISKL 180
Qy 216 KGYHMDPCCSQPEEGVTSPIPIPVGLPVVSHNGATKSIIDYTLERKNEAKACLOSLT 275
Db 181 KGYHMDPCCSQPEEGVTSPIPIPVGLPVVSHNGATKSIIDYTLERKNEAKACLOSLT 240
Qy 276 SRRQISNIKKEKRLTIDLVTDDKMIENHSTSPSCCFCKESFPGPILPHOHERYLCKM 335
Db 241 SRRQISNIKKEKRLTIDLVTDDKMIENHSTSPSCCFCKESFPGPILPHOHERYLCKM 300

```


QY 336 NEEIAVAVLOPHENIYVKNAGVFDVNDKALLLSVLSSEKGLTSPINPYKDHMSVTKAYYAN 395
DB 301 NEEIAVAVLOPHENIYVKNAGVFDVNDKALLLSVLSSEKGLTSPINPYKDHMSVTKAYYAN 360
QY 396 MEPSDELKLSIAVGLPOEFVKEMFEORRKYQYNSNSPSLERSKPLAPNSNPTTDS 455
DB 361 MEPSDELKLSIAVGLPOEFVKEMFEORRKYQYNSNSPSLERSKPLAPNSNPTTDS 420
QY 456 LLPSPVPMDSITSPSIAELHNSVTSCDPLRLTKSSHFTNIIKAVDKLDSRSTPSP 515
DB 421 LLPSPVPMDSITSPSIAELHNSVTSCDPLRLTKSSHFTNIIKAVDKLDSRSTPSP 480
QY 516 NLSSTSSKNSHSSSYTPNSFSSSEELQAEFFRLVITKTERTOKYSHKEONKS 568
DB 481 NLSSTSSKNSHSSSYTPNSFSSSEELQAEFFRLVITKTERTOKYSHKEONKS 533

RESULT 3

US-10-029-386-33350
; Sequence 33350, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AECOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33350
; LENGTH: 533
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010130.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.45
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
; OTHER INFORMATION: SWISSPROT HIT: O60315, EVALU0 0.00e+00
US-10-029-386-33350

Query Match 52.3%; Score 2612; DB 12; Length 533;
Best Local Similarity 94.2%; Pred. No. 4e-145;

Matches 502; Conservative 12; Mismatches 19; Indels 0; Gaps 0;

QY 36 EKPVECPNCKRFSHSGSYSHISSKCGILISVNGRMNNIKTSSPNSVSSSPTNSAI 95
DB 1 EKPVECPNCKRFSHSGSYSHISSKCGILISVNGRMNNIKTSSPNSVSSSPTNSAI 60
QY 96 TOLRKLKENGKPLNSSEOTGLIKITEPLDNDYVLAATHGSSSPMNGGLCATSP 155
DB 61 TOLRKLKENGKPLNSSEOTGLIKITEPLDNDYVLAATHGSSSPMNGGLCATSP 120
QY 156 GVHPSAOSPMOHLGVGMEAPLGLFPTMNSNLSVOVKVLOIVNTYSROMDCKEDISL 215
DB 121 GVHPSAOSPMOHLGVGMEAPLGLFPTMNSNLSVOVKVLOIVNTYSROMDCKEDISL 180
QY 216 KGYHMKDCSOPBEQGVTSNPIPVGLPVSHNGATKSIIDYTLKVNKAACLOSLT 275
DB 181 KGYHMKDCSOPBEQGVTSNPIPVGLPVSHNGATKSIIDYTLKVNKAACLOSLT 240
QY 276 SRRQISNKKKELRTLLDLYTDDKMIENHSISTPSCQPCSKSPFGPIPLHGERYLCM 335
DB 241 SRRQISNKKKELRTLLDLYTDDKMIENHSISTPSCQPCSKSPFGPIPLHGERYLCM 300
QY 336 NEEIAVAVLOPHENIYVKNAGVFDVNDKALLLSVLSSEKGLTSPINPYKDHMSVTKAYYAN 395
DB 301 NEEIAVAVLOPHENIYVKNAGVFDVNDKALLLSVLSSEKGLTSPINPYKDHMSVTKAYYAN 360
QY 396 MEPSDELKLSIAVGLPOEFVKEMFEORRKYQYNSNSPSLERSKPLAPNSNPTTDS 455
DB 361 MEPSDELKLSIAVGLPOEFVKEMFEORRKYQYNSNSPSLERSKPLAPNSNPTTDS 420

DB 361 MEPSDELKLSIAVGLPOEFVKEMFEORRKYQYNSNSPSLERSKPLAPNSNPTTDS 420
QY 456 LLPSPVPMDSITSPSIAELHNSVTSCDPLRLTKSSHFTNIIKAVDKLDSRSTPSP 515
DB 421 LLPSPVPMDSITSPSIAELHNSVTSCDPLRLTKSSHFTNIIKAVDKLDSRSTPSP 480
QY 516 NLSSTSSKNSHSSSYTPNSFSSSEELQAEFFRLVITKTERTOKYSHKEONKS 568
DB 481 NLSSTSSKNSHSSSYTPNSFSSSEELQAEFFRLVITKTERTOKYSHKEONKS 533

RESULT 4

US-09-974-298-114
; Sequence 114, Application US/09974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Huel-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 114
; LENGTH: 1125
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID NO. US20020156263A1 3267030CD1
US-09-974-298-114

Query Match 33.2%; Score 1657.5; DB 10; Length 1125;
Best Local Similarity 41.9%; Pred. No. 8.9e-89;

Matches 406; Conservative 147; Mismatches 308; Indels 107; Gaps 32;

QY 2 LTOGAGNRKFKCTGCAFKYKHLKHELRTHSGEKVEPCNCKRFSHSGSYSHISSK 61
DB 232 VTGSGCNKFKCTGCGRAFKYKHLKHELRTHSGEKVEPCNCKRFSHSGSYSHISSK 291
QY 62 KCIGLISVNGRMNNIKTSSPNSVSSSPTNSAITOLRKLKENGKPLNSSEOTGLIKI 119
DB 292 KCISLIPNKGPRGKLTSSQCSSP-SLSASBSPRPIRQIKER-KL--QEOUSVNOI 347
QY 120 KTEPLDNDYVLAATHGSSSPMNGGLCATSPGLGVHPSAOSPMOHL---GVGMEAPL 176
DB 348 KTEPLDNDYVLAATHGSSSPMNGGLCATSPGLGVHPSAOSPMOHL---GVGMEAPL 176
QY 177 LGFTMNSNLSVOVKVLOI-VDNVYSRQMDCKEDISLKGYNMKDCSOPBEQGVTS 235
DB 407 ---SINLSIDIQVLLKVAADVAVIRQVLE---NNQNLASKEQETINASPLOQGGHSHV 457
QY 236 NIPVGLPVSHNGATKSIIDYTLKVNKAACLOSLTDDSRQISNKKKELRTLLDLY 295
DB 458 -ISAISLPLVODGTTKIINYSLEQSQGVVONLKKRPVATNSCKSEKLPBDDLTVK 516
QY 296 TD-DKMIENH-SISTPSCQPCSKSPFGPI-PLHGERYLCMNEEIAVAVLOPHENIYV 352
DB 517 SEKQSFEGGVNDSTCLLDCDC---PGDINALDELKHY-----DLKQPTQP-----PP 561
QY 353 KAGVFDVNDKALLLSVLSSEKGLTSPINPYKDHMSVTKAYYANMEPSDELKLSIAVGL 412
DB 562 LPAAEAEKPESSVSSATDGNLSPQPLKXLLSLKAYYALNKPSEBELSKTADSVNL 621
QY 413 POEFVKEMFEORRKYQYNSNSPSLERSKPLAPNSNPTTDSLLPSPVPMDSITSPS 472
DB 622 PLDVKKMFEGMOAGQIS-----VQSSBSPBEGKV-----NTPAKNNOPOGSAN 667
QY 473 IAEHNSVTSCDPLRLTKSSHFTNIIKAVDKLDSRSTPSPNLSSTSSKNSHSSSYTP 532
DB 473 IAEHNSVTSCDPLRLTKSSHFTNIIKAVDKLDSRSTPSPNLSSTSSKNSHSSSYTP 532

```

Db 668 ANEPDSTVNLQSLPKMTNS---PVLPGVSTNGSRSTSPSPPLNLSRRNTQGYLYTA 724
Qy 533 NSFSSEELQAEPLDLSPKQREPKGIATKNTKATSIINLDHNSVSSSENS-----DE 587
Db 725 EG-AQEPQVEPVLDSLPRKQGE-----LLERSTTTSVYQNSVYSQVE 767
Qy 588 PLNLTLTKKEFSNNLDKSNPNVFGMNPFAKPLTYLPPOSAPPATFMPVQTSIP 647
Db 768 PLNLSCAKKEPKQKSCVTD--SEPVNVVILPPSANPINIALPTVTAQLPTVAILADQNSVP 825
Qy 648 GLRPYGLDQMSFLPMATYTPPGAATFADMQRKRYQKQFGQGLDLDGADYMSGLD 707
Db 826 CLRALANKQITLIPQVAYTSTTVSP-AVQEPPLKVIQPNQODEQDTSSEGVNSVD 864
Qy 708 MTDSDCLSRKKIKTESGMVACDLCDKTFQKSSLLRHKYETGKRPQCQICKAFFK 767
Db 885 QNDSSTPPKCKKRTKENGWYACDLCKIFQKSSLLRHKYETGKRPHECGICKAFFK 944
Qy 768 KHLIHSRLHSGEKPYQCKGCRFSHSGSYSOHNNHRYSTCKRAEREAERAREK 827
Db 945 KHLIHSRLHSGEKPYQCKGCRFSHSGSYSOHNNHRYSTCKRAERERDSTEOR---- 1000
Qy 828 GHIGPTELLMNRAYLQSTP-OGYSDSERESMPR--DGESEKHEKEGE-----GYK 879
Db 1001 -EAGP-EILSNEHVGAASPSQG--DSDERESLTRREDESEKEEEDKEMELOEKE 1056
Qy 880 LRRRODDEEESESEKSNMDTDEETIRDEEETGDHSMDSSED-----GKMET 930
Db 1057 CEKPGQDEEEGEVEEVEEVE-FAENGEBAKTEGLMKDRAESQASSLGQKVGESSE 1115
Qy 931 KSDHEEDN 938
Db 1116 QVSEKTN 1123

RESULT 5
US-10-104-047-2506
; Sequence 2506, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104, 047
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 2506
; LENGTH: 1104
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2506

```

```

Query Match 33.0%; Score 1647.5; DB 12; Length 1104;
Best Local Similarity 41.8%; Pred. No. 3,4e-88;
Matches 405; Conservative 146; Mismatches 310; Indels 107; Gaps 32;

```

```

Qy 2 LITGAGNRKKKCTECGAFKTKHLEHLIHSGEKPYECPNCKKRPSSHSGSYSHSSK 61
Db 211 VTQSGCNRRKCTECGAFKTKHLEHLIHSGEKPYECPNCKKRPSSHSGSYSHSSK 270
Qy 62 KCIIGLSVNGRMNNIKTG--SSPNVSASPSTNSATQLNKLKENGKLSMSBOTGLTKI 119
Db 271 KCIISLPVNGRPRTGLTKTSCSSP-SISASPSGPTRPQIKOKLEN-KPL--QEQLSVNOI 326
Qy 120 KTEPLDFNDYKVLMAATHGFSGSPFMNGLGATSPGLVHPSAOSPMQHT--GVGMEAPL 176
Db 327 KTEPVDY-BFKPIVVASGINSCTPLQNGVPTGGPGLQATSSPGQWQAVVLPVGLVSP 385
Qy 177 LGPPTMNSNLSVQKYLQI-VDNVTVSQKMDCKTEDISKLGKGYHMDPCSGPEEGYTSP 235

```

```

Db 386 -----SINLSDIQNVLKAVDGNVIRQYLE--NNQANLAKSEQETINASP1QCGSHV 436
Qy 236 NIPVGLPVVSHNGATKSIIDYTLKEVNEAKACLOSLTDSRROISNIKEKRLTIDL 295
Db 437 -ISALSPLVDQDGTTKIINYSLEBPQQLQVVPQULKENPVATNSCKSEKLPEDLYK 495
Qy 296 TD-DKMTENH-SISTPSCQFCESFPPI-PLHQERLYCKMNEIKAVLQPHENIVN 352
Db 496 SEKDQSFEGVNDSTLLCDDC--PGDINALPELKH-----DLKQPTQ-----PP 540
Qy 353 KAGVFDNKAALLSSVLSKGLTSPINPKDHMSYKAYYAMMBEPPSELKISIAVGI 412
Db 541 LPAAEKKESSVSAATGNGNLSPPQPLKNTLSKATYALNAQSAEELSIAVASV 600
Qy 413 POFVKEWFEORKVQYNSRSRPSLERTSKPLAPNSPPTKQSLPRSPVKPMDSTSPS 472
Db 601 PLDVVKKWFEKMQAGIS-----VQSEPESSPEKGV-----NIPAKNNDQQSAN 646
Qy 473 IAEHNSVTSQDPLRLTSSHFTNIKAVDKLDHSRNTSPPLNLSSTSSKSHSSYTP 532
Db 647 ANEPDSTVNLQSLPKMTNS---PVLPGVSTNGSRSTSPSPPLNLSRRNTQGYLYTA 703
Qy 533 NSFSSEELQAEPLDLSPKQREPKGIATKNTKATSIINLDHNSVSSSENS-----DE 587
Db 704 EG-AQEPQVEPVLDSLPRKQGE-----LLERSTTTSVYQNSVYSQVE 746
Qy 588 PLNLTLTKKEFSNNLDKSNPNVFGMNPFAKPLTYLPPOSAPPATFMPVQTSIP 647
Db 747 PLNLSCAKKEPKQKSCVTD--SEPVNVVILPPSANPINIALPTVTAQLPTVAILADQNSVP 804
Qy 648 GLRPYGLDQMSFLPMATYTPPGAATFADMQRKRYQKQFGQGLDLDGADYMSGLD 707
Db 805 CLRALANKQITLIPQVAYTSTTVSP-AVQEPPLKVIQPNQODEQDTSSEGVNSVD 863
Qy 708 MTDSDCLSRKKIKTESGMVACDLCDKTFQKSSLLRHKYETGKRPQCQICKAFFK 767
Db 864 QNDSSTPPKCKKRTKENGWYACDLCKIFQKSSLLRHKYETGKRPHECGICKAFFK 923
Qy 768 KHLIHSRLHSGEKPYQCKGCRFSHSGSYSOHNNHRYSTCKRAEREAERAREK 827
Db 924 KHLIHSRLHSGEKPYQCKGCRFSHSGSYSOHNNHRYSTCKRAERERDSTEOR---- 979
Qy 828 GHIGPTELLMNRAYLQSTP-OGYSDSERESMPR--DGESEKHEKEGE-----GYK 879
Db 980 -EAGP-EILSNEHVGAASPSQG--DSDERESLTRREDESEKEEEDKEMELOEKE 1035
Qy 880 LRRRODDEEESESEKSNMDTDEETIRDEEETGDHSMDSSED-----GKMET 930
Db 1036 CEKPGQDEEEGEVEEVEEVE-FAENGEBAKTEGLMKDRAESQASSLGQKVGESSE 1094
Qy 931 KSDHEEDN 938
Db 1095 QVSEKTN 1102

```

```

RESULT 6
US-10-232-561-10
; Sequence 10, Application US/10232561
; Publication No. US20030119772A1
; GENERAL INFORMATION:
; APPLICANT: Genetex, Thomas
; TITLE OF INVENTION: Methods and compositions useful for
; FILE REFERENCE: CHOP 00-99
; CURRENT APPLICATION NUMBER: US/10/232,561
; PRIOR FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: 60/317,300
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FaesSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 1154
; TYPE: PRT

```

ORGANISM: Human
US-10-232-561-10

Query Match 32.2%; Score 1606.5; DB 15; Length 1154;
Best Local Similarity 41.1%; Pred. No. 9e-86;
Matches 399; Conservative 142; Mismatches 314; Indels 115; Gaps 32;

2 LTQAGNRKFKCTECGKAFKFKYKHLKEHLRIHSGEKPEYECNCKKRFSGSHSSK 61
265 VTQSGNRKFKCTECGKAFKFKYKHLKEHLRIHSGEKPEYECNCKKRFSGSHSSK 324
62 KCIIGLISVNGRMNNIKTG--SSPNSVSSPTNSAITQIRKLENGKPLMSBOQGLKI 119
325 KCIISLIPVNGRPRKLTQSQCSP--SLSPSPSPRPPIRQIKIEN-KPL--QEQLSVNOI 380
120 KTEPLDFNDYKVLMTATGFGSSSPMNGGLATSPGLGHPASQSMOH--GVMGAPL 176
381 KTEPVYD-EFKPIVVASGINSCTPLQNGVFTGGGFLQATSSPQGMVQAVLPTVGLVSP 439
177 LGFPMTNSLSEVQKVLQI-VDNTVSRQKMDCKTEDISKLGKYNMKDPCSQPEQGVTS 235
440 -----SINLSDIQVNLKAVDGNVIRQVLE--NNQANLMSKEQETINASPIQGGHSV 490
236 NIPPVGLVNVSHNGATKSIIDYTELEKVNKACLOSITTSRQISNKKKRLTLIDLV 295
491 -ISALSLPLVDQDGTTKIITINSLRQPSQLQVPPQNLKKNFVATNSCKSEKLPEDLTVK 549
266 TD-DKMIENH-SISFPSCQPCSKSPGPI-PLHQHERYLCMNEIRKAVIOPHENIYPN 352
550 SEKDSFEGGVNDSTCLLCCDC---PDINALPELKYH-----DLQPTP-----PP 594
353 KAGVFNVDKALLLSVLESEKGLTSPINPYKDMSVLKAYYAMNPNSEDLKISAVGL 412
595 LPAEAEKPESSVSSATGDGNLSPSQPLKXNLSTLKKAYALNAPSAEELSKINDSVNL 654
413 POEFYKEMPEQARKYQYNSRSPSLERTSKPLAPNSNPTTKDSLPRSPVKMDSITSPS 472
655 PLDVYKXKPEKMOAQOIS-----VQSESPSSPEPGKV-----NIPAKNNDPOGAN 700
473 IAEHLNSVTSQDPLRLTKSSHFTNIKAVDKLDRSMTPSLNLSNKKSHSSSTTP 532
701 ANEPDSTVNLQSPKMTNS--PVLPGSTTNGSRSSPTSPSLNLSNRTQGYLYTA 757
533 NSFSEELQAEPLDLSLPKQRE-----PKGIATKNKTKATSNILDNHSSSSSENS 585
758 EG-AQEEQVEPLDLSLPKQGELELRIPLPQCLFCGRT----- 797
586 DEPLNLTPIKKEFSNSNNLDNKSNNPFGMNPFSAPLYTLPPLQSAFPAPATFMPVQTS 645
798 ---LELVAKKEPQKOSCVTD--SEPVNVVLPSPANPINIAIPVTQAQLPTVIALDQNS 852
646 IGLRPRYGLQMSLPLHMAVYTPGATFADMOQRKQKQKQSGQGLDQADQYMSGL 705
853 VPCFLALANKTILIPVAYTYSITVSP-AVQEPPLKVIQPNQODERQODTSSGVSANV 911
706 DDMTDSDSLARKIKTESGMVACDLCPKQKSSLLRHYEHTGKRPPOCOICCKAF 765
912 EDQNDSDSTPRKQKRTKENGMTACDLCDKIFQKSSLLRHYEHTGKRPPOCOICCKAF 971
766 KHKHHLIHSRLHSGEKPYQCDKCGKRFSGSHSGSYQHMNHRYSYCKREAEEBAERBAR 825
972 KHKHHLIHSRLHSGEKPYQCDKCGKRFSGSHSGSYQHMNHRYSYCKREAEEBDETEGE-- 1029
826 EKGHGLPELMLNRAVLOSITP-QGYSDESESESPR--DGSEKKEHEGE-----CY 877
1030 ---EAGP-EILSNHVGARASPOG--DSDERESLITREDEBSEKEEBEDKEMELOEE 1083
878 GKLRRDDEEESE 928
1084 KECEKPGQDEEESE 1142
929 ETKSDEEDN 938

DB 1143 SEQVSEKXTN 1152

RESULT 7
US-10-232-561-12
Sequence 12, Application US/10232561
Publication No. US20030119772A1
GENERAL INFORMATION:
APPLICANT: Genetia, Thomas
TITLE OF INVENTION: Methods and compositions useful for
FILE REFERENCE: CHOP 00-99
CURRENT FILING DATE: 2002-08-30
PRIOR APPLICATION NUMBER: 60/317,300
PRIOR FILING DATE: 2001-09-05
NUMBER OF SEQ ID NOS: 12
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 12
LENGTH: 1117
TYPE: PRT
ORGANISM: Mouse
US-10-232-561-12

Query Match 31.4%; Score 1566; DB 15; Length 1117;
Best Local Similarity 39.7%; Pred. No. 2e-83;
Matches 393; Conservative 142; Mismatches 310; Indels 144; Gaps 32;

2 LTQAGNRKFKCTECGKAFKFKYKHLKEHLRIHSGEKPEYECNCKKRFSGSHSSK 61
210 VTQSGNRKFKCTECGKAFKFKYKHLKEHLRIHSGEKPEYECNCKKRFSGSHSSK 269
62 KCIIGLISVNGRMNNIKTG--SSPNSVSSPTNSAITQIRKLENGKPLMSBOQGLKI 119
270 KCIISLIPVNGRPRKLTQSQCSP--SLSPSPSPRPPIRQIKIEN-KPL--QEQLSVNOI 326
120 KTEPLDFNDYKVLMTATGFGSSSPMNGGLATSPGLGHPASQSMOHLAGVM 172
327 KTEPVYD-EFKPIVVASGINSCTPLQNGVFTGGGFLQATSSPQGMVQAVLPTVGLVSP 381
173 EAPLGFPTMNSNLSEVQKVLQI-VDNTVSRQKMDCKTEDISKLGKYNMKDPCSQPEQGV 231
382 VSP-----SINLSDIQVNLKAVDGNVIRQVLETNQ--ASLASKQEVASAPLOOG 432
232 VTSNIPPVGLVNVSHNGATKSIIDYTELEKVNKACLOSITTSRQISNKKKRLTL 291
433 GHSV-ISALSLPLVDQDGTTKIITINSLRQPSQLQVPPQNLKKNFVATNSCKSEKLPED 491
292 IDLVTD-DKMIENH-SISFPSCQPCSKSPGPI-PLHQHERYLCMNEIRKAVIOPHENI 349
492 LTVASETDKSEFGARDDSTCLCEDC-----PGDL-----NALPELKHVDPECPA 537
350 VPKAGVFNVDKALLLSVLESEKGLTSPIN-PYQDHMSVLKAYYAMNPNSEDLKISAVGL 408
538 QPPPPAPATEKPESSAS--AGNDLSPSQPLKXNLSTLKKAYALNAPSAEELSKINDSVNL 595
409 AVGLQAEVEMPEQARKYQYNSRSPS-----LEKTSKPLAPNSNPTTKDSL 456
596 VSNLPLDGVKXKPEKMOAQOIP--QSPDPSPGTSVNIPTKTBQPPALGAGNPQEDST 654
457 LPRSPVKMDSITSPSLAEHLNSVTSQDPLRLTKSSHFTNIKAVDKLDRSMTPSLN 516
655 RGQSPVKIRSTPVP-VGSANNGSRSC-----TSSPSPLN 688
517 LSTSSKNSHSSSTPNPSFSEELQAEPLDLSLPKQREPKGIATKNKTKATSNILDNH 576
689 LCSARNPGYS--CVAGAQEEQVEPLDLSLPKQGE-----LTERS 729
577 SVSSSENS-----DEPLNLTPIKKEFSNSNNLDNKSNNPFGMNPFSAPLYTLPPLQPS 631
730 TVSSVYQNSYSVOELPLNLSCAKKEPQKOSCVTD--SEPVNVVLPSPANPINIAIPVT 787
632 APPPATFMPVQTSIIGLRPRYGLDQMSFLPHMAVYTPGATFADMOQR--KYQKQSG 689

```

Db      788 ACPFTVAIADQNSVPLRLAANKQTILPOVAAY---SATVPAVDEPPVKYIQPNG 844
Qy      690 FCGDLLDGAQDYVSGJDDMTDSCLSRKKIKKTESGMYACDLCDTPOKSSSLHKKXE 749
Db      845 NQDERQDTSSEGGTYVEDQNDSDSTPKKTKRTKTEGMWACDLCDDIPIKSSSLHKKXE 904
Qy      750 HTGKRPHQOICCKAFKFKHHLIHSRLHSGEKPYQCDKCGKRFSGSGSYQHMHNRYSY 809
Db      905 HTGKRPHQOICCKAFKFKHHLIHSRLHSGEKPYQCDKCGKRFSGSGSYQHMHNRYSY 964
Qy      810 CREAEERAEAREAEKHLGPTL--LMNRAVLOSITPGYSDSESESMR--DGS 865
Db      965 CKRGADRDAMEQE---DAGEVLPVYLATEHVGAARASPSQADSDERESLTREEDDS 1019
Qy      866 EKEHEKEGE---GYKTLRRDGDSEEESEENKSMDDTPETIRDEETGDH---- 917
Db      1020 EKEHEEDKEMELQSGKEENQSGHEEESEEESEEESEEEVEVADENHMAATDGT 1079
Qy      918 -----SMDSEEDGKMTKSDHEE 936
Db      1080 VEVGAQAQAGSLERQKASESESESESEQ 1108

```

```

RESULT 8
US-09-864-761-34862
; Sequence 34862, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/224,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203

```

```

; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34862
; LENGTH: 190
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010130.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 6.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6.8
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 6.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.7
; OTHER INFORMATION: EST_HUMAN HIT: BF245591.1, EVALUATE 1.00e-53
; OTHER INFORMATION: SWISSPROT HIT: P37275, EVALUATE 2.00e-33
US-09-864-761-34862

```

```

Query Match      20.1%; Score 1005; DB 9; Length 190;
Best Local Similarity 97.4%; Pred. No. 1.6e-51;
Matches 185; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy      753 KRPHQOICCKAFKFKHHLIHSRLHSGEKPYQCDKCGKRFSGSGSYQHMHNRYSYCKR 812
Db      1 KRPHQOICCKAFKFKHHLIHSRLHSGEKPYQCDKCGKRFSGSGSYQHMHNRYSYCKR 60
Qy      813 EAERERAEAREAREKHLGPTL--LMNRAVLOSITPGYSDSESESMR--DGESEKEHEKE 872
Db      61 EAERERAEAREAREKHLGPTL--LMNRAVLOSITPGYSDSESESMR--DGESEKEHEKE 120
Qy      873 GEGYCKLRRDGDSEEESEENKSMDDTPETIRDEETGDHSMDSSEDKMTKTS 932
Db      121 GEGYCKLRRDGDSEEESEENKSMDDTPETIRDEETGDHSMDSSEDKMTKTS 180
Qy      933 DHEEDNMDG 942
Db      181 DHEEDNMDG 190

RESULT 9
US-09-864-761-36988
; Sequence 36988, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664

```



```

385  ISQSSSSDDUDVSTPSGSDLEFY-----SSDLESDIESDKEKFXENGKMFK 432
QY 454 DILLPSPYKPM-----STPSIAE--LHNSTSCDPRRLTKSHFNIAVD 502
Db 433 DKSPLONLASINNKKEYNHSIFSPSLEOTAIVAAGVANDSIKAIASIAEKYFGSTGLV 4922
QY 503 KLDHSRNTPSPLNTLSTSSKNSHSSSYPNFSSEELQAEPLDI-----SLPKOM 5533
Db 493 LQDKYKALPYR-SMPLFPFPFQSYM--FPDRDIRSLPLNKPQSPBEYVLQKGS 5499
QY 554 REBKGIATKNTKATSIINLDHNSVSSSSSEDEPLNTFTIKGEFNSNNLDKNNPVF 6133
Db 550 SESPPDLTKRKOEXELTPPEKRPVTPATSDQDLDSMGSRSPASGTYKTEPRKNHVF 6099
QY 614 G-----MNPSPAPLYTLPLPQSAFPAPATFMPV-----QTSIP-----GLRPY 6522
Db 610 GKKKGSVBSRPSDGL-----QIARTPFPMPIYRVERKKTDLLEALKERYLRPS 6633
QY 653 PGL---DOMSFLPHMAVYTPGTGATPADMQQRKQORQKQFQGLDLDAQOYMGLDMT 7099
Db 664 PGLFHPQNSAINMA-----EKLESFSA-LKPASLLOSVSMF 7033
QY 710 D---SDSCLRKKIKKTESGMYACDLCDXTFOKSSLLRHKYEHTKRPHOCQICKAFF 7666
Db 704 NFRAPRALPENLIRKQKE-RYTCRCYCKIFPSANLTRHRTHTGEQPYRCYCDRFS 7622
QY 767 HKHHLLEHSR-LHSGERPYQCDKCGKRFSSHSSYSQHM-NRYSICRKAEREAARREA 8244
Db 763 ISSNLQRHVANINHKKEPRKCHLCRQFOQTNLRLHKKHENGMSGTATVSSPSELES 8222
QY 825 -----REKHLGP--TELLMRAVLYOS-----ITPOGYS--DSE 8544
Db 823 TGA1LDDKEDAYTELRNFIQNSNHCOSQPRAYVERBMMGSHFKDEKALVYSQNSDLLDE 8822
QY 855 ERESMPRDSGEKEHEKEGBEGYGL--RRRDGEDEEEEBEESENKSMIDTPETIRDEE 9122
Db 883 EVEDEVLLDEDEDNDITGKTGKEPVTSLHAGNEDDYEEFSALEMSCKTSPVRYKEE 9422
QY 913 -ETGDHSM-----DSEDEKKEKTSHEED 937
Db 943 YKSGSLADHTRHFTDSLKRRKEDNOYSEAE 974

```

```

RESULT 11
US-10-097-340-79
Sequence 79, Application US/10097340
Publication No. US20030087250A1
GENERAL INFORMATION:
APPLICANT: John MONAHAN
APPLICANT: Manjula GANNAVARAPU
APPLICANT: Sebastian HOERSCH
APPLICANT: Shubhangi KAMATKAR
APPLICANT: Steve G. KOVATS
APPLICANT: Rachel E. MEYERS
APPLICANT: Michael MORRISSEY
APPLICANT: Peter OLANDT
APPLICANT: Ami SEN
APPLICANT: Peter VEIBY
APPLICANT: Gordon B. MILLS
APPLICANT: Robert C. BAST, Jr.
APPLICANT: Karen LU
APPLICANT: Rosemarie SCHMANDT
APPLICANT: Xumel ZHAO
APPLICANT: Karen GLATT
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
Title Of Invention: Assessment, Prevention, and Therapy of Ovarian Cancer
FILE REFERENCE: MRI-030
CURRENT APPLICATION NUMBER: US/10/097.340
CURRENT FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: 60/276,025
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/325,149
PRIOR FILING DATE: 2001-09-26

```

```

? PRIOR APPLICATION NUMBER: 60/276,026
? PRIOR FILING DATE: 2001-03-14
? PRIOR APPLICATION NUMBER: 60/324,967
? PRIOR FILING DATE: 2001/09/26
? PRIOR APPLICATION NUMBER: 60/311,732
? PRIOR FILING DATE: 2001-08-10
? PRIOR APPLICATION NUMBER: 60/325,102
? PRIOR FILING DATE: 2001-09-26
? PRIOR APPLICATION NUMBER: 60/323,580
? PRIOR FILING DATE: 2001-09-19
? NUMBER OF SEQ ID NOS: 363
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 79
? LENGTH: 1051
? TYPE: PRT
? ORGANISM: Homo sapiens
US-10-097-340-79

       7.1%; Score 355.5; DB 15; Length 1051;
Query Match      19.4%; Pred. No. 1,66-12;
Best Local Similarity
Matches 199; Conservative 140; Mismatches 355; Indels 293; Gaps 36

```

Qy	9	RKPCJTCGGAFFKXJHILKEHLYJHSEKPEJECNCKRFFHSHSJSYSHISXKCIJGIS	68
Db	101	REYCDCCGPPAAFNKXJLIRHOSHDSBKATIECNCAKVFETDPBNLQRHITSQHV	155
Qy	69	VNGRMRNNKITGSSPNSSVSSPTNSAITOLANKLENGKPL	121
Db	156	GABAHACPECGKITPATISSGKLKHJHISVYKPFICEVCHKSUYTOFQSNLCRHKR	208
Qy	122	EPPLD	167
Db	209	MEADCRTOIKCKDCGOMFSTTSSLNKRRRPFCEGNHFAAGF	250
Qy	168	LGWGEARPLIGFTPMNSNLSVQKVLQIDVDTVRQKMDCKTEDISKLYGHMKDPCSQP	227
Db	251	FGQGISLP	286
Qy	228	EEQGVTSNP	273
Db	287	HPAGLTPTPAAGFSFVPGLFPSGLYHRPLIPASSPVKGL	338
Qy	274	TDSRRQISNKKKXLTILDLVTDKMIENHSISTPSCQPKESFPQPLIQAHERLJC	333
Db	339		348
Qy	334	KMNEIKAVLOPHENIVPKAGVVDNKKALLSSVLSKGLTSPINPKDMSVLKAYYA	393
Db	349	ATQOILTA	384
Qy	394	MNMEPSNDELKJISIAVGLPOEFVKWEQORQVYOYSNRSRSLERTSKPLAPNSPTTK	453
Db	385	ISDQSESDLDVSTPSGSDLETT	432
Qy	454	DSLIPASPVKPM	502
Db	433	DKVSPLONLASINNKKKEYSNHSIFSPSLEQTAVASGAVNOSIKAIASIAEKYFGSTGLVG	492
Qy	503	KLHRSRNTSPPLNLSSTSKNSHSSSYTPSPFSEELQAEPLD	553
Db	493	LQDKKVGALPVP	549
Qy	554	REPKGIIATKUKTKATSIINLDHNSVSSSESDSEPLNTFIKKEFFSNMNLNDKNSNNPVF	613
Db	550	SESEFDLTTKRKDEKPLTPVPSKRPPVTPATSDQDPLDMSRBRASOTKLTERRKHVF	609
Qy	614	G	652
Db	610	GKKKGSNVESRPADGSL	663
Qy	653	PGLDQMSPLPMAATYTPGATPADMOQRKRYQKQSGQSLDGAQDYMSGLDMDTP	710
Db	664	PG	715

Qy	711	-SDSCSRKKIKKTEGMACDLCKTFOKSSLLRHKYBHTGKRPBHCOCCKAFAFGHK	769
Db	716	APPNALPENLILKGMK-RYTCRYCGKIPPRASNLRLHRLTGTGQPRCKCDRSFESS	774
Qy	770	HLIEHSR-LHSGKPYOCDKCGKRPSSGSGYSQHMNHRYSYCKRAEEREAEREAREK	828
Db	775	NQHRNRNLNHNKKPKFCHLCYRCGQOOTNLDRH-----KKIENG	815
Qy	829	HLGPTLLNRAVLOSITPGYSDSEERESMPDSESEKEHKE---GEGYGLRLRDG	885
Db	816	NMSGT-----ATSSPHSELSETCALILDDEDAFPFEIRNFIGNSHSGQSRRNV	864
Qy	886	DE-----EEBEEBESFNKSM---DTDPETLRDEEETGDHMSDSSDEGKMETYSDH	934
Db	865	EEEMNGSHKFEKKAALVPQNSDILLDDEVEVEVLDEE---DEDYITGKTGKEPVTYSL	921
Qy	935	EEEDNMD	941
Db	922	HEGNPFD	928

```

RESULT 12
US-10-029-386-33060
; Sequence 33060, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ. ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33060
; LENGTH: 1204
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC003973.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.7
; OTHER INFORMATION: SWISSPROT HIT: O43345, EVALU 0.00e+00
; US-10-029-386-33060

```

[illegible]

QY	208	----	KTEDISXKGYMADPCSOEEOGYVTSFPIPIVGLPVSINNGATKSIIDYLTLEKVN	264
Db	709	AFNBSAILIKKRIHITDEKPYKCEEGKTFKSVSTLTKAHAG-----	EKPVK	758
QY	265	AKACLOSLTSDSRROISNIKKCKLRTLIDLVTDDKMIENHSISTPSCQKESFPGRIP	324	
Db	759	CCEGGAFSKFS-----	ILTKHKVI--HTGEKPYKCECGKAYKMPST	799
QY	325	LHOHR-----	YICK-----MNEIKAVLOPHENIVPNKAGVVD--NRKLLSSVLSK	372
Db	800	LSYHKKHITGEKPYKCECGKGFMSFSLTKHVVHITGEKPYKCECGKAFSMLSVFSKH	859	
QY	373	GLTSPINPYKDHMSVLYAYAMNMPNSDELKISIAVLPOEFVKEWFEQKRYOYSNS	432	
Db	860	KKTIAGEKPYKCECGKAY-----	KSSSTL-----SYHKKHITGEKPYK--	899
QY	433	RSPSLERTSKPLAPNSNPTTKDLSLPRSPVKEMDSTSPSIAELHNSVTS	CDPPLRTKS	492
Db	900	-----	EECGK-----GPFESILTKHKVVIHTEKPYKCECG	930
QY	493	SHFTNIKAVDKLDHSRNT-----	PSPLNLSSTSSKNSHSSSYTPNSFS	536
Db	931	GKAFNWS--NLMEKKIHTGETPYKCEBDKAFSMPSSL-----	TEHKATHAGE--KPYK	982
QY	537	SEELQAEPLDLSLPKOMEPKGIATKNTKATSIINLDHNSVSSSENSEDEPLNLTPIKK	596	
Db	983	CEEGCKA--FSMPRLTEHK-----	ATHAGEPYKCECGK	1016
QY	597	EFSNSNNLDKNSNNVPGMNPSPAKPLTYPLPQ-----	SAPPAATFMPVQISIP	647
Db	1017	AFNWSNN-----	MEHRIHTEKPYKCECGKSPSTSLTKHKVVIH--	1060
QY	648	GLRYPGLDQMSFLPHMAVYPTGATPFADMOQRK--YORKOGFOCDLIDGAODYMSGLD	706	
Db	1061	GEKPYK--CEEGK-----	KAYKMSSTLSYHKKHITVEKPYKCECGKPFV-----	1104
QY	707	DMTSDGCLSRKKTIKTESGMATCDLCTFQKSSSLNKKYEHGTGRPHQOQICKAKFK	766	
Db	1105	-----	FSLIAKHKVIHTEKPYKCECGKAYKMPSTLRYHKKHITGEKPYKCECGKAFS	1159
QY	767	HKHHLEHSRLSHSGEPYOCDCGKRPFSHSGYSOH	802	
Db	1160	TFSLITKAKVIHTEKPYKCECGKAFSMLSVFSKA	1195	

```

1 RESULT 13
2 US-09-864-761-38036
3 Sequence 38036, Application US/09864761
4 Patent No. US20020048763A1
5 GENERAL INFORMATION:
6 APPLICANT: Penn, Sharon G.
7 APPLICANT: Rank, David R.
8 APPLICANT: Hanzel, David K.
9 APPLICANT: Chen, Wensheng
10 TITLE OF INVENTION: HUMAN GENOME- DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
11 TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
12 FILE REFERENCE: Aemica-X-1
13 CURRENT APPLICATION NUMBER: US/09/864,761
14 CURRENT FILING DATE: 2001-05-23
15 PRIOR APPLICATION NUMBER: US 60/180,312
16 PRIOR FILING DATE: 2000-02-04
17 PRIOR APPLICATION NUMBER: US 60/207,456
18 PRIOR FILING DATE: 2000-05-26
19 PRIOR APPLICATION NUMBER: US 09/632,366
20 PRIOR FILING DATE: 2000-08-03
21 PRIOR APPLICATION NUMBER: GB 24263.6
22 PRIOR FILING DATE: 2000-10-04
23 PRIOR APPLICATION NUMBER: US 60/236,359
24 PRIOR FILING DATE: 2000-09-27
25 PRIOR APPLICATION NUMBER: PCT/US01/00666
26 PRIOR FILING DATE: 2001-01-30
27 PRIOR APPLICATION NUMBER: PCT/US01/00667

```



```

1 PRIOR FILING DATE: 2001-01-30
2 PRIOR APPLICATION NUMBER: PCT/US01/00664
3 PRIOR FILING DATE: 2001-01-30
4 PRIOR APPLICATION NUMBER: PCT/US01/00669
5 PRIOR FILING DATE: 2001-01-30
6 PRIOR APPLICATION NUMBER: PCT/US01/00665
7 PRIOR FILING DATE: 2001-01-30
8 PRIOR APPLICATION NUMBER: PCT/US01/00668
9 PRIOR FILING DATE: 2001-01-30
10 PRIOR APPLICATION NUMBER: PCT/US01/00663
11 PRIOR FILING DATE: 2001-01-30
12 PRIOR APPLICATION NUMBER: PCT/US01/00662
13 PRIOR FILING DATE: 2001-01-30
14 PRIOR APPLICATION NUMBER: PCT/US01/00661
15 PRIOR FILING DATE: 2001-01-30
16 PRIOR APPLICATION NUMBER: PCT/US01/00670
17 PRIOR FILING DATE: 2001-01-30
18 PRIOR APPLICATION NUMBER: US 60/234,687
19 PRIOR FILING DATE: 2000-09-21
20 PRIOR APPLICATION NUMBER: US 09/608,408
21 PRIOR FILING DATE: 2000-06-30
22 PRIOR APPLICATION NUMBER: US 09/774,203
23 PRIOR FILING DATE: 2001-01-29
24 NUMBER OF SEQ ID NOS: 49117
25 SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
26 SEQ ID NO 38036
27 LENGTH: 269
28 TYPE: PRT
29 ORGANISM: Homo sapiens
30 FEATURE:
31 OTHER INFORMATION: MAP TO AC005877.2
32 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6
33 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
34 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
35 OTHER INFORMATION: EST HUMAN HIT: BE046326.1, EVALU8 9.00e-68
36 OTHER INFORMATION: SWISSPROT HIT: P37275, EVALU8 1.00e-116
37 JS-09-864-761-38036

```



```

; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1199
; LENGTH: 1357
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-1199

```

```

Query Match      6.4%; Score 318; DB 12; Length 1357;
Best Local Similarity 18.7%; Pred. No. 3.5e-10;
Matches 163; Conservative 94; Mismatches 235; Indels 380; Gaps 28;

```

```

QY 7 GNRKCTECGKAFYKHLKEHLIHSGEKPECPNCKRPFSGSYSHISSKICGL 66
DB 693 GEKPYRCECGKAFWPSNLTRHKRIHTEGKPYACEGCGAFRRSSTLTNH----- 743
QY 67 ISVNGRMNNITGSSPNSVSSPTNSAITQARNLENKGPLSMSEQTGLKITEPLDF 126
DB 744 -----KRIHTGERP----- 752
QY 127 NDYKYLMTATHGSPGSGSPFMNGLGATSPGVHPSAQSFMQHLGVGMEAPLGLPTMNSNL 186
DB 753 --YKCEBGCKAFSVSSALI----- 769
QY 187 SEVQKVLQIVDNTVSRQKDKCTEDI SKLKGVMKDPGQPEQGVTSBNIPVGLPVVS 246
DB 770 -----YHKR----- 774
QY 247 HNGATKSIIDYTLKENVKAKACLOSLTDSRROISNIKEKRLIDLVDTKMIENHSI 306
DB 775 HTGER---PYTCEBGKAFNCSSTLKT-----HKII--HTG 805
QY 307 STPFCQFCCKESFPPIPLHOHER-----YLCKMNERIKAVLOPHENIVPNKAGVFD 359
DB 806 EKPYTCEBGKRTFNCSSTYKAKHRIHTGEKPYK--EBCDAFKMH----- 849
QY 360 NKALLLSVLSKGLTSPINPYKDHMSVLKAYYAMNMBNSDELKISIAVGLPOEFYKE 419
DB 850 -----SLAKKHIIHTGEKPYK-----SDSKALAKSS----- 877
QY 420 WFEQKRYQ---YSNSRSPSLERTSKPLAPNSNPPTTKOSLPRSPYKPMDSITSPSIAEL 476
DB 878 --EVQKVVSGDENGIRVHKKETQGWLVRRNNENRTGLFOIRAAVRPN--RDPSSWQO 932
QY 477 HNSVTSQDPPLTLTK---SSHFTNIKAVDK-----LDHS 507
DB 933 EGSLLT--DPIQKEEPDLQNHIDHONALEDOQNTGVGILLTRDVVTEPSLEEWQCLDHA 990
QY 508 RSNTPSPNLSTSSKNSHSSSYT--PNSFSEELQAEPLDL-----SLPKOM--R 554
DB 991 QONLVRDVLNENYRVLVSLGIAVSKPDLITCLEQNKPEPNNIRGNEMVTGHPDLPELGIK 1050
QY 555 EPKGIATNKTAKAISINDHNSVSSSE-----NSDEPLNLTFLK----- 595
DB 1051 DSLQKVIPIRRYKSGHDNLQVYTKCSMGCEVQKGCNEVNOCLSTQNKIPQTHKCVKV 1110
QY 596 -KEFSNNSNLDKSNPNVFGMNPFSAPKPLVTLPPQSAFPATFMPVPVQTSIRGLRYPYG 654
DB 1111 RGFPSNSNR--HKTHT--GKHGFKCKY----- 1135
QY 655 LDQMSFLPHMAYTYPTGATFADMQORRYQ---RKQFGQDLLDGAQDYNSGLDMDTD 710

```

```

DB 1136 -----GKSFQWVSQLOHQIHTRENSYQCE--EGKPF-----N 1168
QY 711 SDSCISRRKIKKTESGMVACDLCDKTFOKSSSLRHKYEHTGKAPHOQICKAFKHH 770
DB 1169 CSSTLSKHKRIHTEGKPYRCECGKAFWSSSTLTNRKRIHTEGKPYTCEBGCGAFPSNST 1228
QY 771 LIEHSRLHSGEKPYQCDKCGRRFSGSYSQH 802
DB 1229 LANHKRIHTEGKPYTCEBGKAFSLSSSLTVH 1260

```

Search completed: January 5, 2004, 16:37:15
Job time : 56 secs

THIS PAGE BLANK (USPTO)